

Diseño de la Arquitectura de Referencia para la Gestión de Datos Maestros en Colombia – Versión 3.0

Fase 3

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



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 Saenz - Apoyo administrativo/Supervisor Técnico

MinTIC

Luis Climaco 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	17/02/2023	Creación de documento base Desarrollo Actividad 1	Marta Silvia Tabares Sandra Liliana Santos
2.0	13/03/2023	Desarrollo completo de las 3 actividades de la Fase 3	Marta Silvia Tabares Sandra Liliana Santos
3.0	10/04/2023	Ajustes observaciones finales	Marta Silvia Tabares Pedro F. Pérez Arteaga

Contenido

Control de cambios	3
Contenido	4
1. Objetivo General y Específicos del Diseño de la Arquitectura de Referencia para la Gestión de Datos Maestros	9
2. Metodología de Trabajo	10
3. Marco Teórico	11
3.1. Autenticación Digital	11
3.2. Autenticación Multifactor	11
3.3. Interoperabilidad Gubernamental	11
3.4. Carpeta ciudadana	14
3.5. Resolución de Entidad	14
4. Soluciones Tecnológicas	15
4.1. Características de Evaluación de la Solución Tecnológica	15
4.1.1. Tipo de Concentrador GDM	15
4.1.2. Arquitectura Técnica	16
4.1.3. Usabilidad	17
4.1.4. Gestión de Versiones para los datos	18
4.1.5. Seguridad y Privacidad de los datos	18
4.1.6. Flujos de Trabajo	20
4.1.7. Gobernanza de los Datos	20
4.1.8. Resolución de Entidad	21
4.1.9. Sincronización	22
4.1.10. Enfoque de Gestión de Datos	23
4.1.11. Gestión de Monitoreo, alertas y notificaciones	23
4.1.12. Condiciones de entrega	23

4.2. Alternativas de Solución Tecnológica para la Gestión de Datos Maestros	24
5. Diseño de la Arquitectura de Referencia	27
5.1. Componentes de la Arquitectura de Referencia	27
5.1.1. Entidades Fuentes y Consumidoras	27
5.1.2. Plataforma de Interoperabilidad – X-ROAD	28
5.1.3. Zonas para la Gestión de Datos maestros y de referencia	29
5.1.4. Componentes de apoyo a la gestión	31
5.2. Ciclo de Vida de los Datos Maestros y de Referencia	36
5.2.1. Ciclo de vida de los datos maestros	36
5.2.2. Ciclo de vida de los datos de referencia	43
5.3. Análisis de consideraciones especiales y recomendaciones	45
6. Modelo de Datos del Concentrador GDM	46
6.1. Auditoria de Cambios	47
6.2. Tablas Administrativas	48
6.2.1. Modelo de datos históricos	48
6.2.2. Modelo del Catálogo de Datos	52
6.3. Modelo de Datos – Tablas Maestras y de Referencia	53
6.1. Tablas Maestras	54
6.4. Tablas de Referencia	63
7. Análisis de Soluciones para la Gestión de Datos Maestros	71
Recomendaciones para el análisis de soluciones para la Gestión de Datos Maestros	75
Glosario	7
8	
Referencias	8
1	
Anexo 1. Dimensiones y Reglas de Calidad de los Datos	88
8.1.1. Dimensión de Exactitud	88
8.1.2. Dimensión de Completitud	89
8.1.3. Dimensión de Consistencia	90
8.1.4. Dimensión de Unicidad	91

8.1.5. Dimensión de Validez	91
8.1.6. Dimensión de Oportunidad	92
8.1.7. Dimensión de Vigencia	92
8.1.8. Dimensión de Conformidad	93
8.1.9. Dimensión de Integridad	93
8.1.10. Dimensión de Precisión	93
8.2. Reglas de emparejamiento (matching).....	94
8.3. Reglas de Supervivencia	95
Anexo 2. Proceso Estrategia y Gobierno para la GDM Nivel Estratégico	96
Anexo 3. Proceso de Seguridad y Protección de los datos	99

Lista de Figuras

FIGURA 1. CUADRANTE MÁGICO DE GARTNER – SOLUCIONES PARA LA GESTIÓN DE DATOS MAESTROS - AÑO 2022 VERSUS “THE FORRESTER WAVE VENDOR REVIEW” AÑO 2021.	25
FIGURA 2. ARQUITECTURA DE REFERENCIA PARA LA GESTIÓN DE DATOS MAESTROS Y DE REFERENCIA EN EL CONCENTRADOR GDM. FUENTE: ELABORACIÓN PROPIA.	28
FIGURA 3. PROCESO EL CICLO DE VIDA DE LA GESTIÓN DE DATOS MAESTROS. ELABORACIÓN PROPIA.	37
FIGURA 4. SUBPROCESO: EXTRAER DATOS. FUENTE: ELABORACIÓN PROPIA.	38
FIGURA 5. SUBPROCESO: INTEGRAR DATOS. FUENTE: ELABORACIÓN PROPIA.	41
FIGURA 6. CICLO DE VIDA DE LOS DATOS DE REFERENCIA.	44
FIGURA 9. MODELO DE DATOS DEL CONCENTRADOR GDM. FUENTE: ELABORACIÓN PROPIA.	46
FIGURA 10. MODELO DE DATOS HISTÓRICO DE CAMBIOS.	49
FIGURA 11. MODELO DE DATOS DE LAS TABLAS ADMINISTRATIVAS QUE SON PARTE DEL CATÁLOGO DE DATOS. FUENTE: ELABORACIÓN PROPIA.	52
FIGURA 12. MODELO DE DATOS – TABLAS MAESTRAS Y DE REFERENCIA. FUENTE: ELABORACIÓN PROPIA.	54
FIGURA 13. MODELO DE DATOS DE LA TABLA MAESTRA DE PERSONAS HUBDPER_MAESTRAPERSONA – DOMINIO DE PERSONAS.	55
FIGURA 14. MODELO DE DATOS DE LA ENTIDAD MAESTRA DE RUES EN EL ROR – DOMINIO DE EMPRESAS.	58
FIGURA 15. MODELO DE DATOS DE LA MAESTRA DE CATASTRO HUBDLUG_MAESTRACASTRO – DOMINIO LUGARES.	60
FIGURA 16. MODELO DE REFERENCIA DE DIRECCIÓN – HUBREF_DIRECCION.	64
FIGURA 17. MODELO DE REFERENCIA DE USUARIO – TABLA HUBREF_USUARIO.	66
FIGURA 18. MODELO DE REFERENCIA – TABLA HUB_REFERENCIA.	68
FIGURA 17. ARQUITECTURA DE REFERENCIA DE INFORMATICA MDM (HTTPS://DOCS.INFORMATICA.COM/MASTER-DATA-MANAGEMENT/MULTIDOMAIN-MDM/10-3/OVERVIEW-GUIDE/INFORMATICA-MDM-HUB-ARCHITECTURE.HTML).	72
FIGURA 18. ARQUITECTURA DE REFERENCIA MDM – RELTIO (HTTPS://WWW.RELTIO.COM/MASTER-DATA-MANAGEMENT/).	73
FIGURA 19. ARQUITECTURA DE REFERENCIA MDM – IBM CLOUD (HTTPS://WWW.IBM.COM/DOCS/EN/IMDM/12.0?TOPIC=OVERVIEW-ARCHITECTURE-CONCEPTS).	74
FIGURA 20. JERARQUÍA PARA EL ASEGURAMIENTO DE LA CALIDAD DE LOS DATOS MAESTROS Y DE REFERENCIA.	88
FIGURA 21. PROCESO ESTRATEGIA Y GOBIERNO PARA LA GDM NIVEL ESTRATÉGICO.	96
FIGURA 22. PROCESO GESTIÓN Y GOBIERNO DE DATOS NIVEL TÁCTICO.	97
FIGURA 23. PROCESO SEGURIDAD Y PROTECCIÓN DE DATOS.	99

Lista de Tablas

TABLA 1. ACTIVIDADES DEFINIDAS PARA EL DESARROLLO DE LA FASE 3. (FUENTE: ELABORACIÓN PROPIA)	10
TABLA 2. SERVICIOS DE INTEROPERABILIDAD TÉCNICA DEFINIDOS POR (NACER, 2021).	12
TABLA 3. PROCESO DE INTERCAMBIO DE DATOS (UNA ADAPTACIÓN DE (MINVIVIENDA, 2021)).	29
TABLA 4. DESCRIPCIÓN DE PERMISOS PARA EL CONCENTRADOR GDM.	34
TABLA 5. PERMISOS PARA LOS GRUPOS DE ROLES DEFINIDOS EN EL MARCO DE GOBIERNO DEL CONCENTRADOR GDM.	34
TABLA 6. PERMISOS POR USUARIO DEL CONCENTRADOR GDM.	35
TABLA 7. CARACTERÍSTICAS DEL REGISTRO DE ORO O DATOS DE REFERENCIA DE ORIGEN (ROO).	39
TABLA 8. TABLAS DEL MODELO DEL CONCENTRADOR GDM	47
TABLA 9. DICCIONARIO DE DATOS DE LA TABLA HUBADM_HISTORICOCAMBIOS.	49
TABLA 10. DICCIONARIO DE DATOS MAESTRA DE PERSONA – HUBDPER_MAESTRAPERSONA.	56
TABLA 11. DICCIONARIO DE DATOS MAESTRA DE PERSONA – HUBDPER_UBICACIONPER	57
TABLA 12. DICCIONARIO DE DATOS DE LA TABLA MAESTRA DEL RUES HUBDEPM_MAESTRARUES – DOMINIO EMPRESAS.	58
TABLA 13. DICCIONARIO DE DATOS DE LA TABLA MAESTRA DE CATASTRO HUBDLUG_MAESTRACATASTRO – DOMINIO LUGARES.	61
TABLA 14. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA DE DIRECCIÓN HUBREF_DIRECCION.	64
TABLA 15. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_DIVIPOLA.	65
TABLA 16. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_DEPARTAMENTO.	66
TABLA 17. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_COMUNA.	66
TABLA 18. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUBREF_USUARIO.	67
TABLA 19. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_REFERENCIA.	68
TABLA 20. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_TIPOREFERENCIA.	69
TABLA 21. DICCIONARIO DE DATOS DE LA TABLA DE REFERENCIA HUB_TIPOREFERENCIA.	70
TABLA 22. CARACTERÍSTICAS DE MERCADO DE LA SOLUCIONES MDM EN NUBE (LAS MEJORES CALIFICADAS POR GARTNER PEER INSIGHTS).	71
TABLA 23. EJEMPLO DE CATEGORÍAS Y TIPOS DE LICENCIAMIENTOS.	75

1. Objetivo General y Específicos del Diseño de la Arquitectura de Referencia para la Gestión de Datos Maestros

1.1. Objetivo General

Diseñar una arquitectura de referencia que responda a lo definido en el diseño del modelo de gestión de datos maestros para el Estado colombiano, y que sea aplicable a una solución de distribución tecnológica que será desarrollada en una etapa posterior.

1.2. Objetivos Específicos

1. Revisión de referentes sobre modelos, mecanismos y soluciones tecnológicas y de arquitecturas para su implementación que apliquen para el caso colombiano. ([capítulo 4](#))
2. Elaboración del documento técnico que incluya el diseño de la arquitectura de referencia que responda a lo definido en el diseño del modelo de gestión. Este documento debe incluir los modelos de alto nivel agnósticos a la tecnología, y restricciones y recomendaciones para su implementación. ([capítulo 5](#) y [capítulo 6](#))
3. Elaborar un análisis de alternativas de solución para la distribución de datos maestros que incluya la pertinencia de integración con X-ROAD (Plataforma de interoperabilidad del Estado colombiano). ([capítulo 7](#))

2. Metodología de Trabajo

La metodología para lograr el objetivo específico de la consultoría, “OE3: Diseñar una arquitectura de referencia que responda a lo definido en el diseño del modelo de gestión de datos maestros para el Estado colombiano”, está definida a partir de las actividades, que se describen en la Tabla 1.

Tabla 1. Actividades definidas para el desarrollo de la Fase 3. (Fuente: Elaboración Propia)

Actividad	Tareas	Responsables	Entregable
1. Identificar soluciones tecnológicas	<ul style="list-style-type: none"> Identificar las características que permitan evaluar posibles soluciones tecnológicas para la implementación de la gestión de datos maestros del Gobierno de Colombia Seleccionar tres soluciones tecnológicas que cumplan con las características identificadas y que sean relevantes dentro del mercado tecnológico a nivel mundial 	Consultoría CAOBA Apoyo: Equipo DNP-MinTIC	Documento entregable Fase 3 – Entrega final
2. Diseñar la Arquitectura de Referencia	<ul style="list-style-type: none"> Definir la arquitectura que sea referente para la implementación del modelo de gestión de datos maestros definido en la Fase 2. Describir los componentes de la arquitectura y la forma como ellos interactúan. Generar la versión final de los modelos de datos, el ciclo de vida de los datos, el catálogo de datos, y del aseguramiento de calidad definidos en la Fase 2 	Consultoría CAOBA	Documento Entregable Fase 3 – Entrega final
3. Realizar un análisis de capacidades de las soluciones tecnológicas seleccionadas	<ul style="list-style-type: none"> Evaluar componentes y capacidades de las soluciones tecnológicas seleccionadas Recomendar los elementos mínimos que deberán ser tenidos en cuenta para dar continuidad a la implementación de la arquitectura de referencia propuesta 	Consultoría CAOBA Apoyo: Equipo DNP-MinTIC	Documento Entregable Fase 3 – Entrega Final

3. Marco Teórico

En esta sección se presentan algunos de los conceptos tecnológicos más relevantes que son usados en diferentes capítulos de este documento.

3.1. Autenticación Digital

(MinTIC - SCD, 2020) define la autenticación digital como “el procedimiento que permite verificar los atributos digitales de una persona cuando adelanten trámites y servicios a través de medios digitales en las entidades, mediante el cual se puede afirmar que dicha persona es quien dice ser. Para el acceso a este servicio dichas entidades deben identificar y determinar el nivel de garantía a partir del grado de confianza requerido para sus procesos, siendo estos niveles de confianza clasificados como bajo, medio, alto y muy alto.”

3.2. Autenticación Multifactor

“Es un método de autenticación electrónica en el que se otorga acceso a un usuario a un sitio web o una aplicación solo después de presentar con éxito dos o más pruebas a un mecanismo de autenticación: conocimiento, posesión e inherencia.” (ComputerWeekly.es, 2023).

3.3. Interoperabilidad Gubernamental

“La interoperabilidad gubernamental es un requisito para hacer posible la comunicación digital y el intercambio de información entre las administraciones públicas, y entre estas y las empresas privadas y los organismos no gubernamentales que deban interactuar con el Estado, con la finalidad de lograr un mercado digital único.” (Nacer, 2021).

(MinTIC - SCD, 2020) define la interoperabilidad, en el marco de servicios ciudadanos digitales, como: “Capacidad de las organizaciones para intercambiar información y conocimiento en el marco de sus procesos de negocio para interactuar hacia objetivos mutuamente beneficiosos, con el propósito de facilitar la entrega de servicios en línea a ciudadanos, empresas y a otras entidades, mediante el intercambio de datos entre sus sistemas.”

Servicios de Interoperabilidad Técnica. (Nacer, 2021) define el servicio de interoperabilidad técnico como el “servicio que se encarga de definir y promover arquitecturas tecnológicas para la interoperabilidad de cada institución para cada servicio que es provisto/consumido, mediante la gestión de un catálogo de servicios, un catálogo de esquemas y metadatos, un directorio de datos y la administración de un registro de trazabilidad y monitoreo.” La tabla 2 muestra nueve servicios de interoperabilidad técnica propuestos por (Nacer, 2021).

Tabla 2. Servicios de Interoperabilidad técnica definidos por (Nacer, 2021).

Nro.	Servicio	Responsabilidad de la Institución de interoperabilidad	Responsabilidad de otras Instituciones
1	Suministro y administración de un gestor de convenios	Proveer y administrar un gestor de convenios (aplicación) global y específico entre instituciones	Mantener actualizada la aplicación que facilita la tramitación de convenios electrónicos estándar, entre consumidor y proveedor de servicios de interoperabilidad, entregando los permisos y las credenciales de acceso para autorizar el suministro del servicio
2	Suministro y administración de un registro de trazabilidad	Proveer y administrar un registro de trazabilidad	Mantener actualizado el registro resultante de cada servicio de interoperabilidad para que los actores involucrados conozcan y verifiquen las operaciones que se hayan realizado respecto de un determinado servicio interoperado
3	Suministro y administración de un sistema de monitoreo	Proveer y administrar un sistema de monitoreo	Mantener actualizado el sistema de monitoreo que realiza consultas periódicas a todos los servicios de interoperabilidad definidos en el catálogo de servicios, a fin de identificar y notificar de manera oportuna y proactiva las fallas en el servicio
4	Suministro y administración de un gestor de autorizaciones	Proveer y administrar un gestor de autorizaciones que permita a las personas naturales autorizar o desautorizar el uso de sus datos personales, mediante el uso de su identidad digital	N/A
5	Identificación de los servicios adecuados para que se invoquen y organicen para prestar el servicio público	Garantizar que las necesidades se identifiquen y que los servicios adecuados se invoquen y organicen para prestar un servicio público	Formalizar los servicios demandados por la ciudadanía y las interdependencias que estos tienen con otras instituciones

Nro.	Servicio	Responsabilidad de la Institución de interoperabilidad	Responsabilidad de otras Instituciones
6	Suministro y gestión de un catálogo de fuentes de información y servicios internos de cada institución	Proveer y gestionar un catálogo de fuentes de información (aplicaciones) de cada institución	Publicar y mantener en el catálogo el listado de sistemas y aplicaciones desde donde se obtienen los servicios y documentos electrónicos interoperables
7	Suministro y gestión de un catálogo de recursos reutilizables	Proveer y gestionar un catálogo de recursos reutilizables de interoperabilidad entre instituciones	Publicar y mantener en el catálogo de recursos reutilizables los módulos, componentes y servicios que puedan reutilizar otras instituciones
8	Suministro y gestión de un catálogo de fuentes de información y servicios externos	Proveer y gestionar un catálogo de fuentes de información y servicios externos utilizados por cada institución	Publicar y mantener el catálogo de fuentes de información y servicios externos, identificando el proveedor y documentando las características técnicas y el uso de estas fuentes
9	Suministro y administración de un mecanismo de seguridad y privacidad de acceso a la información	Proveer y administrar un mecanismo de seguridad y privacidad de acceso a la información interoperable	Hacer uso de las funcionalidades y los ambientes del mecanismo de seguridad y privacidad de acceso a la información interoperable

(MinTIC - SCD, 2020) define “Servicio de Interoperabilidad: es el servicio que brinda las capacidades necesarias para garantizar el adecuado flujo de información e interacción entre los sistemas de información de las entidades, permitiendo el intercambio, la integración y la compartición de la información, con el propósito de facilitar el ejercicio de sus funciones constitucionales y legales, acorde con los lineamientos del marco de interoperabilidad.”

(MinTIC - Marco de interoperabilidad, 2019) específicamente define que “el dominio técnico de la interoperabilidad hace referencia a las aplicaciones e infraestructuras que conectan sistemas de información, a través de los servicios de intercambio de información. Incluye aspectos como especificaciones de interfaz, protocolos de interconexión, servicios de integración de datos, presentación e intercambio de datos y protocolos

de comunicación seguros. La interoperabilidad técnica debe garantizarse, siempre que sea posible, mediante el uso de especificaciones técnicas formales y de los Servicios Ciudadanos Digitales.”

La gestión de los datos maestros del Gobierno de Colombia define la interoperabilidad de datos en la arquitectura de referencia propuesta, como el mecanismo requerido para interactuar con las entidades participantes, proveedoras o consumidoras (ver principio 8 y política 18 definidos en documento Fase 2),

3.4. Carpeta ciudadana

(MinTIC - SCD, 2020) define la carpeta ciudadana como el “servicio que le permite a las personas naturales o jurídicas acceder y gestionar digitalmente de manera segura, confiable y actualizada al conjunto de sus datos, que tienen o custodian las entidades y solicitar actualización de esto.”

3.5. Resolución de Entidad

Es el proceso que reconocer una entidad maestra específica y la representa adecuadamente de manera única, completa y precisa desde el punto de vista de los datos. La resolución de entidades es importante porque aborda el problema de las múltiples versiones de la verdad. Abarca procesos y técnicas para identificar y resolver las muchas ocurrencias de una entidad maestra a través de muchas fuentes.

4. Soluciones Tecnológicas

En esta sección se definen un conjunto de características que deben guiar la evaluación de posibles soluciones tecnológicas para la implementación de la arquitectura de referencia definida para la gestión de datos maestros del Gobierno Nacional. Las características permitirán orientar las etapas futuras de implementación definidas en los pasos 2, 3, y 4 propuestos en la hoja de ruta presentada en el documento de la Fase 2. Además, se presentan los proveedores de soluciones tecnológicas para la gestión de datos maestros, mejor calificados en el mercado de tecnologías, con las correspondientes características de evaluación definidas por *Gartner Peer Insights* y *The Forrester Wave*.

4.1. Características de Evaluación de la Solución Tecnológica

Las características de evaluación se agrupan en doce áreas clave¹:

- Tipo de concentrador GDM
- Arquitectura técnica
- Usabilidad
- Gestión de versiones
- Seguridad y Privacidad de los datos
- Flujos de trabajo
- Gobernanza de Datos
- Resolución de Entidad
- Sincronización
- Enfoque de gestión de datos
- Gestión de Monitoreo, alertas y notificaciones
- Condiciones de entrega

4.1.1. Tipo de Concentrador GDM

Para la implementación del Concentrador para la Gestión de los Datos Maestros (Concentrador GDM), la solución tecnológica debe evaluarse desde dos posibilidades: La primera, basada en un repositorio virtual que utilice un índice maestro que vincule, mediante un apuntador, a los registros originales en sus conjuntos de datos de origen, donde obliga a la creación de un solo registro maestro y permite que diferentes entidades consumidoras vean registros maestros usando diferentes reglas de presentación. La otra alternativa, basada en un repositorio que posibilite la convivencia de un solo "registro de oro" o registro consolidado que selecciona valores de atributos que sobreviven entre los registros vinculados a una misma entidad maestra o de referencia definida en

¹ La definición de las áreas es presentada a partir de la experiencia de la consultoría y adaptada a algunos componentes propuestos por referentes como IBM, Microsoft, etc. en términos de soluciones tecnológicas.

el modelo de datos del Concentrador GDM. Este registro se convierte en un activo de consulta que las entidades consumidoras acceden usando una o diferentes reglas de presentación.

4.1.2. Arquitectura Técnica

Los Administradores de datos maestros requieren una aplicación web o plataforma en la nube que provea escenarios de colaboración fáciles de instalar, mantener e implementar. Esto requiere una arquitectura simple, un cliente de huella cero² y fácil acceso para los usuarios del Concentrador GDM. La plataforma tecnológica de la base de datos del *back-end*³ debe ser invisible para los usuarios finales. Además, esta la solución tecnológica debe proveer funciones para la configuración y el mantenimiento de los diferentes componentes definidos en la arquitectura de referencia. La solución tecnológica debe permitir almacenar en varios esquemas o instancia de base de datos, proporcionar manejo de dominios agrupados por diferentes estilos de implementación de la arquitectura, y proveer componentes y servicios para facilitar acciones de Administrador o acciones relacionadas a los custodios u operadores de los datos tales como gestionar, instalar y mantener la usabilidad general de la aplicación (p.ej., aplicar pruebas de fuego), cuando sea requerido. Tanto la arquitectura lógica como física deben definir componentes asociados al encapsulamiento, la escalabilidad y la flexibilidad.

- **Encapsulamiento.** La solución tecnológica deberá proveer la capacidad de desacoplar la información de las aplicaciones para la gestión de los datos. Es decir, la arquitectura debe hacer la gestión de los datos independiente de los servicios de datos que sean configurados “bajo demanda”, y que a su vez sean accedidos por diferentes tipos de usuario en el momento y el contexto el cual la aplicación fue diseñada.
- **Escalabilidad.** La solución tecnológica debe proveer una fuente centralizada y autorizada de datos maestros del Gobierno de Colombia que gestiona la integridad de los datos proporcionados por las fuentes y controla la distribución de datos maestros en las diferentes entidades destino de forma estandarizada. Esto facilita la reutilización y la escalabilidad.
- **Flexibilidad.** La solución tecnológica debe proveer la flexibilidad necesaria para adaptarse a los cambios en los requisitos y regulaciones definidas por el Gobierno de Colombia y admisión de nuevos dominios y datos maestros.
- **Licenciamiento en la nube.** En la actualidad, los tipos de licenciamiento más comunes de computación en la nube son: licencias a demanda (*on-demand*), licencias de pago por uso (*pay-as-per-use*), y licencias de corto alcance (*short-range*). Para adquirir estos tipos de licenciamiento, los vendedores de nube ofrecen varios modelos de licencia (**Buyya, Vecchiola, & Selvi, 2013**):
 - **Modelo de usuarios concurrentes.** El cliente compra un conjunto de licencias, donde el *check-in* y *check-out* de la licencia está asociado a una sesión web o la forma como está definido el

² Un cliente de huella cero significa que no hay instalación o descarga del lado del cliente, lo que permite la visualización de documentos e imágenes dentro del navegador web nativo, mientras aprovecha las verdaderas capacidades del navegador, los complementos integrados y su interacción con el propio dispositivo.

³ Capa de acceso a datos (back-end) de una pieza de software. o la infraestructura física o el hardware.

funcionamiento de una aplicación, ya sea por módulos, opciones de menú, o APIs. Generalmente, se aplica a través de enlaces (hooks) de aplicación que invocan a un servicio de administrador de licencias. En el servicio de nube multi-cliente, se puede subcategorizar el grupo de licencias para dar asignaciones a los clientes, que pueden no ser reforzadas por el software de administración de licencias.

- **Modelo del titular de los derechos de autor.** En este modelo, los principales proveedores de servicios en la nube suelen preferir una combinación de software gratuito o de código abierto y software de proyectos de desarrollo nacionales. Las empresas que poseen los derechos de autor, ya sea sobre la construcción o la compra, pueden usar el software para cualquier propósito que deseen.
- **Modelo de usuario designado.** En este modelo, una licencia está limitada a un cliente en particular. Ese cliente tiene licencia para implementar el software en cualquier tipo de dispositivo y en una cantidad de dispositivos simultáneamente. Por lo general, las licencias están vinculadas al directorio de servicios internos de la organización, pero los servicios basados en la nube utilizan proveedores de identidad de Internet. Las licencias de usuario nominal se intercambian con una organización con un grupo muchas veces, que luego se pueden asignar y migrar según las necesidades.

Los beneficios o restricciones de cada tipo de licenciamiento con respecto a la escalabilidad o crecimiento se deben definir de forma clara y ajustados a las necesidades particulares que se definen desde la arquitectura de referencia, Con respecto al escalamiento, este se debe poder balancear con los tipos de licencia y modelo que se adquiera ya sea en volumen de datos y tiempos de ejecución de alguna aplicación específica.

Se sugiere no pactar acuerdos o contratos a largo plazo con respecto a los volúmenes y el período de tiempo.

- **Generador de reportes e informes.** La arquitectura técnica debe incluir un componente para la generación de reportes e informes, el cual ponga a disposición una serie de información administrativa acorde a las necesidades de los diferentes roles definidos en el Concentrador GDM. Los informes producidos por el sistema deben estar disponibles en todo momento en una variedad de formatos, incluidos HTML, PDF, Excel, archivo web, TIFF, texto y XML. Además, debe facilitar la generación de informes ad-hoc y personalizados de fácil acceso sobre cualquier dato maestro dentro del sistema teniendo en cuenta permisos de acceso a los datos definidos en el componente de seguridad.

4.1.3. Usabilidad

La usabilidad es una de las características de calidad de las soluciones de software más importantes al momento de elegir una solución tecnológica. Según la ISO 25000 (ISO, 2023), la usabilidad se refiere a la “capacidad del software para ser entendido, aprendido, usado y atractivo para el usuario, cuando se usa bajo determinadas condiciones. Esta característica se subdivide a su vez en las siguientes sub características:

- **Reconocimiento de la adecuación.** Capacidad del producto que permite al usuario identificar si el software es adecuado para sus necesidades.
- **Aprendizaje.** Capacidad del producto que permite al usuario aprender su uso y aplicación.
- **Operatividad.** Capacidad del producto que permite al usuario operarlo y controlarlo con facilidad.
- **Protección contra errores de usuario.** Capacidad del sistema de aplicar medidas para prevenir, detectar y corregir errores cometidos por los usuarios.
- **Estética de la interfaz de usuario.** Capacidad de la interfaz de usuario de agrandar y satisfacer la interacción con el usuario.
- **Accesibilidad.** Capacidad del producto que permite que sea utilizado por usuarios con determinadas características de restricciones de acceso.”

La definición de la primera versión del registro de oro en cada uno de los dominios del sistema de gestión de datos maestros orientará la identificación de capacidades de usabilidad para realizar operaciones CRUD y desarrollar flujos de trabajo ETL. Es importante que el sistema proporcione interfaces de fácil consulta/acceso o facilite el desarrollo de interfaces de usuario final para gestionar los datos maestros o de referencia. Del mismo modo, la solución tecnológica debe proporcionar los mecanismos u opciones para importar datos desde almacenes de datos o sistemas gestores de datos maestros, a través de operaciones SQL, o archivos en formato TXT, XML o JSON. En general, la usabilidad definida por las soluciones tecnológicas evaluadas debe ser una experiencia de usuario que impulse la aceptabilidad de uso desde los diferentes roles que sean definidos en el Concentrador GDM.

4.1.4. Gestión de Versiones para los datos

Versionar datos significa crear una referencia única de una colección de datos en una fecha y hora determinada con una versión de metadatos. Esta referencia puede tomar la forma una ID o un identificador de fecha y hora. La solución tecnológica debe proveer un sistema gestor de versiones relacionada a los datos maestros y de referencia almacenados en el Concentrador GDM. Los metadatos y datos pueden cambiar a través del tiempo por diferentes razones, por eso es necesario que los administradores u otros roles puedan validar datos, revisar y revertir transacciones, crear copias de datos con versiones y marcar versiones de datos.

4.1.5. Seguridad y Privacidad de los datos

Una solución tecnológica para la gestión de datos maestros debe proveer un componente para definir las políticas de seguridad basadas en roles como base para definir los derechos de acceso a cualquier información. El componente debe involucrar funcionalidades como las que se listan a continuación.

- **Autenticación de usuario.** La autenticación es el proceso de verificar la identidad de un usuario. Un usuario de una solución para la gestión de datos maestros puede ser desde un administrador de datos hasta otro usuario registrado para consultar datos o acceder al Concentrador GDM. Estos usuarios se autentican en función de credenciales como un nombre de usuario y una contraseña. La solución

tecnológica debe incluir un protocolo de administración de autenticación de usuarios internos y externos al Concentrador GDM de tal forma que se garantice que los usuarios sean válidos antes de proporcionar acceso a los datos.

- **Autorización de usuario.** La autorización es el proceso de determinar si un usuario tiene privilegios suficientes para acceder a los datos solicitados del gestor de datos maestros. La capacidad de autorización de usuarios dentro de una solución tecnológica le permite proteger varios recursos en el Concentrador GDM permitiendo que solo los usuarios con los privilegios obtengan autorización para acceder a los datos. Esto se puede lograr creando autorizaciones de usuario basadas en roles y el acceso a los recursos que se necesitan para completar las tareas permitidas por el usuario. Desde otro punto de vista, la solución tecnológica debe ofrecer la capacidad de cifrar los atributos de los datos con claves criptográficas definidas por la organización, de modo que a los usuarios se les solicite una contraseña antes de que puedan ver los valores de los atributos.
- **Registros de acceso y marcas de tiempo.** La solución tecnológica para la gestión de datos maestros debe permitir hacer seguimiento de auditoría y capacidades de marca de tiempo. Esto es, ver quién solicitó o accedió a qué registros del Concentrador GDM, cuándo y dónde. Cada acción sobre los datos debe generar alertas y flujos de trabajo para señalar comportamientos inapropiados. La solución tecnológica también debe marcar la hora de cada creación y actualización de atributos, y permitir la publicación de la afectación y aplicación de acuerdos de nivel de servicio (SLA) de acuerdo con los datos afectados y el impacto en la toma de decisiones de las entidades afectadas.
- **Evaluación, corrección y supervisión de los riesgos de datos confidenciales.** La solución tecnológica debe proveer los mecanismos, interfaces, y reportes que faciliten a diferentes roles definidos, analizar, remediar y monitorear los riesgos de los datos confidenciales que se utilizarán en una implementación del en el Concentrador GDM. Esto implica que la solución tecnológica habilite espacios para:
 - Comprender dónde residen los activos de datos confidenciales en las entidades fuente, el Concentrador GDM y los entornos de las entidades consumidoras. La solución tecnológica de proveer mecanismos de descubrimiento y análisis avanzados para determinar la sensibilidad, la ubicación, la función y la proliferación de los datos.
 - Analizar y priorizar el riesgo de datos confidenciales. La solución tecnológica debe proveer calificaciones de riesgo automatizadas para determinar el riesgo basadas en políticas o reglas de negocio definidas por el gobierno nacional. Esto es, regulaciones relevantes asociadas a cómo y quién usa los datos maestros, y también cómo se distribuyen los datos.
 - Remediar el riesgo con controles de seguridad de datos. La solución tecnológica debe proporcionar mecanismos de encriptación, enmascaramiento, tokenización y controles de

acceso para garantizar que solo aquellos con una designación de necesidad de saber puedan ver los datos y, en general, limitar el acceso a los datos por parte de un solo rol.

- Supervisar el acceso a los datos. La solución tecnológica debe proporcionar mecanismos para monitorear factores como accesos excesivos a los datos del Concentrador GDM, el movimiento de datos entre entidades y el Concentrador GDM, alertas sobre políticas implementadas en el concentrador GDM y control a usuarios que buscan acceso excesivo o inusual.
- Realizar proceso de rastreo integral. Esto garantiza que los datos nuevos, los almacenes de datos o sistemas gestores de datos fuente, y los usuarios sean analizados y rastreados en busca de condiciones de riesgo.
- **Integración de capacidades de ciberseguridad.** La solución tecnológica debe proveer los mecanismos para que el equipo de seguridad del Concentrador GDM determine qué controles a nivel de VPN (red privada virtual), firewalls y SEIM (*Security information and event management*) deben activar o implementar para soportar ataques cibernéticos o malware en el entorno del Concentrador GDM.

4.1.6. Flujos de Trabajo

Los flujos de trabajo para la gestión de datos maestros representan actividades específicas o los pasos de trabajo que se requieren para realizar los diversos procesos relacionados con la gestión de datos maestros. La solución tecnológica de proveer la interfaces y mecanismos requeridos para definir y automatizar trabajos asociados al intercambio de datos entre entidades proveedoras/consumidoras y el Concentrador GDM; proceso que garanticen el ciclo de vida de los datos, sincronización, integración, seguridad, etc. De esta forma, los usuarios administradores o custodios de los datos podrán monitorear, recibir notificaciones automáticas por correo electrónico o mensajería con alertas de eventos clave de gestión de datos maestros.

4.1.7. Gobernanza de los Datos

Cualquier solución tecnológica para la gestión de datos maestros debe tener como base el componente de gobierno de datos operativos y para la analítica. Esta debe tener en cuenta la capacidad de los administradores de datos para realizar funciones tales como: crear roles y responsabilidades a diferentes grupos de usuarios de gobierno y gestión de datos maestros y de referencia, anular decisiones automatizadas para vincular (o no vincular) registros o ajustes a los modelos de datos o metadatos, definir reglas de gobierno de datos asociadas a la seguridad y privacidad de los datos, y gestionar reglas de calidad asociadas a los registros maestros y de referencia.

4.1.8. Resolución de Entidad

“La integración de datos también conocido como resolución de entidades es un componente crítico para permitir la interpretación e identificación de registros a través de múltiples fuentes que están asociadas con la misma entidad. Este proceso implica la identificación de muchas ocurrencias de datos maestros en sistemas dispares; limpiándolos, estandarizándolos, combinándolos y enriqueciéndolos; y finalmente terminar con una sola versión, que a menudo se llama el disco de oro.” (Allen & Cervo, 2015). A partir de la resolución de entidad, el registro de oro se atribuye a la única versión de la verdad para la ocurrencia de una entidad determinada. Así, se puede resolver la ambigüedad que puede darse cuando un registro de oro proviene de diferentes fuentes.

Por esta razón, la solución tecnológica, además de proveer reglas para la limpieza, normalización, y estandarización de los datos, debe proveer los mecanismos necesarios para resolver la entidad o el registro de oro al poder vincular registros similares que provienen de diferentes fuentes. Esos mecanismos pueden emplear tanto coincidencias (emparejamiento) deterministas como coincidencias probabilísticas⁴ (i.e., puntuación de similitud para vincular registros que no comparten exactamente los mismos valores) (Alsarkhi & Talburt, 2018). También es importante que la solución tecnológica permita evaluar la precisión y la exactitud del emparejamiento; es decir, la solución tecnológica debe prescribir límites predecibles de coincidencias de falsos positivos y evitar las no coincidencias de falsos negativos. En este sentido, la solución debe facilitar a los usuarios mecanismos para ajustar la configuración y modificar, por ejemplo, la puntuación de la coincidencia, o crear reglas a la medida. A continuación, se resaltan las funciones de emparejamiento y vinculación, enriquecimiento de los datos, y reglas de negocio.

Emparejamiento y Vinculación de Datos Maestros. Además de reunir diversas fuentes de datos, una solución de gestión de datos maestros debe contener una función de emparejamiento y vinculación. Esto debido a que se pueden presentar eventos de registros o datos duplicados y cambios en la información desde diferentes fuentes. Una capacidad de emparejamiento y vinculación debe utilizar algoritmos que identifiquen automáticamente los duplicados y ayuden a resolver múltiples entradas en un registro único y preciso. La función de emparejamiento y vinculación de datos maestros deberá garantizar lo siguiente, para el Concentrador GDM:

- Eliminación de datos duplicados
- Ingesta de información precisa en todos sus sistemas
- Monitoreo de la integridad de sus sistemas fuente
- Definición de reglas para enriquecer sus registros con diferentes fuentes de datos
- Automatización de tareas que requieran realizarse de forma repetitiva y con poca o ninguna intervención humana.

⁴ Los enfoques determinista y probabilístico son enfoques de emparejamiento comúnmente utilizados en los sistemas de resolución de entidades (ER) en los gestores de datos maestros.

Refinamiento de Registros con Enriquecimiento de Datos. El enriquecimiento de datos se refiere a las herramientas y procesos que mejoran la calidad de los datos provenientes de las diferentes fuentes de entrada. La limpieza y la optimización de los datos ayudan a identificar datos incompletos, incorrectos o irrelevantes y modificarlos o eliminarlos cuando sea necesario para brindar una visión más precisa de los datos almacenados en especial en las entidades de referencia. Una vez los datos han pasado por el proceso de limpieza, se pueden enriquecer al combinarlos con datos provenientes de las fuentes para ampliar aún más su descripción.

Aplicación de Reglas de Negocio de Datos Maestros. Una solución técnica para la gestión de datos maestros debe poder introducir condiciones o acciones que modifiquen los datos en consecuencia de estas. Estas reglas también se definen desde el negocio para determinar la periodicidad de extracción/carga de datos o reglas para se realiza procesos de aprobación de alguna acción determinada sobre los datos. Este tipo de reglas ejecutarse desde el Concentrador GDM para que su mantenimiento se realice una vez y se aplique en todas partes que se requiera. Con la debida seguridad, las reglas de negocio se deben implementar directamente en la interfaz de usuario para no consumir recurso del back-end. En otras palabras, la solución tecnológica debe proporcionar la administración de las reglas de negocio, y así facilitar la creación, aplicación y administración de políticas de gobierno de datos en el Concentrador GDM. Las políticas que se definan para gestionar las reglas de negocio afectarán directamente el ciclo de vida de los datos, los flujos de trabajo para los procesos de decisión y aprobación y las reglas que determinan la integridad de los datos.

4.1.9. Sincronización

La solución tecnológica debe proveer los mecanismos necesarios para que los modelos de datos de las entidades proveedoras y consumidoras se sincronicen con y desde la estructura de datos maestra definida en el Concentrador GDM. Así, los datos de origen deben llevarse al entorno del Concentrador GDM para la resolución, vinculación y consolidación de identidades de forma periódica (por ejemplo, todas las noches). Esta tarea de sincronización se debe configurar con períodos de tiempo exactos para que las aplicaciones de las fuentes no acumulen datos de entidad entre esos períodos; así, los sistemas de datos de las fuentes podrán reconocer los datos 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. En otras palabras, “una vez que el registro de oro se crea, se mantiene y está disponible en el Concentrador GDM, se puede publicar o sincronizar con otras fuentes, según se requiera. Los datos se publican o sincronizan según el estilo de arquitectura. En un estilo registro, hay una única versión virtual de los datos maestros; por lo tanto, la publicación solo es requerida por los sistemas consumidores. En un estilo híbrido, se copian algunos atributos; en consecuencia, la sincronización es necesaria.” (Allen & Cervo, 2015).

4.1.10. Enfoque de Gestión de Datos

Los sistemas de administración de bases de datos relacionales (RDBMS) siguen siendo la alternativa más usada para almacenar datos maestros. Sin embargo, debido a que las conexiones entre los registros de entidades contienen información (incluida la información de identificación), algunos proveedores están reemplazando el RDBMS con bases de datos basadas en grafos para capturar datos maestros. Las bases de datos de grafos tratan los vínculos entre entidades como objetos de primera clase, con sus propios atributos, y proporcionan un medio alternativo para buscar, vincular y analizar datos maestros.

4.1.11. Gestión de Monitoreo, alertas y notificaciones

La solución tecnológica debe proveer un componente para vigilar la existencia y la magnitud del cambio de estado y el flujo de datos en un sistema. “El monitoreo tiene como objetivo identificar fallas y asistir en su subsiguiente eliminación. Las técnicas utilizadas en el monitoreo de los sistemas de información se cruzan con los campos del procesamiento en tiempo real, las estadísticas y el análisis de datos. Un conjunto de componentes de software utilizados para la recopilación de datos, su procesamiento y presentación se denomina sistema de monitoreo. Alertar es la capacidad de un sistema de monitoreo para detectar y notificar a los operadores sobre eventos significativos que denotan un cambio de estado grave. La notificación se conoce como alerta y es un mensaje simple que puede tomar varias formas: correo electrónico, SMS, mensaje instantáneo (IM) o una llamada telefónica. La alerta se transmite al destinatario adecuado, es decir, una parte responsable de tratar el evento. La alerta a menudo se registra en forma de ticket en un Sistema de seguimiento de problemas (ITS), también conocido simplemente como sistema de emisión de tickets. (Ligus, 2012).

4.1.12. Condiciones de entrega

El proveedor de la solución tecnológica contratada debe proporcionar los siguientes entregables, los cuales ayudarán al equipo de gobierno del Concentrador GDM a tener criterios para recibir y poner en marcha la solución propuesta.

- Manuales completos técnicos y de usuario final en formato digital o papel
- Versiones de certificación y puesta en producción de todos y cada uno de los componentes y elementos tecnológicos contratados.
- Demostración de las pruebas realizadas a nivel de plataforma tecnológica, gobierno, seguridad y protección, linaje de datos y toda la funcionalidad contratada para toda la implementación y puesta en marcha del Concentrador GDM
- Planes de actualización de productos, servicios, y procesos de control de calidad documentados, que demuestren la función y la competencia de gestión de los productos o servicios establecidos por contrato.
- Modelo de soporte de producto y servicios contratados incluyendo el detalle del proceso de resolución de problemas

4.2. Alternativas de Solución Tecnológica para la Gestión de Datos Maestros

La identificación de posibles soluciones para la gestión de datos maestros se realiza teniendo en cuenta el cuadrante mágico “la voz del consumidor” de *Gartner Peer Insights* del año 2022 de Soluciones MDM (Master Data Management) (Pimcore, 2023) y el análisis del “*Forrester Wave Vendor Review*” del año 2021 (TIBCO, 2023). Es importante anotar que, al momento de la consulta Forrester no había publicado la versión del año 2022. Aun así, no se descarta porque del año 2021 al año 2022 los vendedores de soluciones tecnológicas no tuvieron cambios significativos, La figura 1 muestra las clasificaciones que otorgan cada una de estas compañías.

Interpretación del análisis hecho por Gartner en (Gartner Peer Insights, 2023) – Figura 1 (a). Gartner en el cuadrante mágico de estas soluciones. En este, el eje X representa «Integridad de la visión», y el eje Y representa «Habilidad para ejecutar».

- **Cuadrante superior derecho: Elección de los clientes**

“*Gartner Peer Insights* reconoce a los proveedores que cumplen o superan tanto la calificación general promedio del mercado como el puntaje promedio de interés y adopción del usuario a través de la distinción *Customers’ Choice*. Los proveedores en este cuadrante recibieron calificaciones sólidas por su experiencia general en relación con el mercado. Este es un buen indicador de la satisfacción con los productos del proveedor en este mercado. Los proveedores en este cuadrante también superan el interés y la adopción de usuarios promedio del mercado, que incluye el volumen de reseñas, la cobertura del mercado de reseñas y la voluntad de recomendar.”

- **Cuadrante superior izquierdo: Fuerte desempeño**

“Proveedores que tienen una calificación igual o superior a la calificación general promedio del mercado, pero que no alcanzan el puntaje promedio de adopción e interés del usuario. Los proveedores en este cuadrante recibieron calificaciones sólidas por su experiencia general en relación con el mercado. Este es un buen indicador de la satisfacción con los productos del proveedor en este mercado. Sin embargo, los proveedores en este cuadrante están por debajo del interés y la adopción del usuario promedio del mercado.”

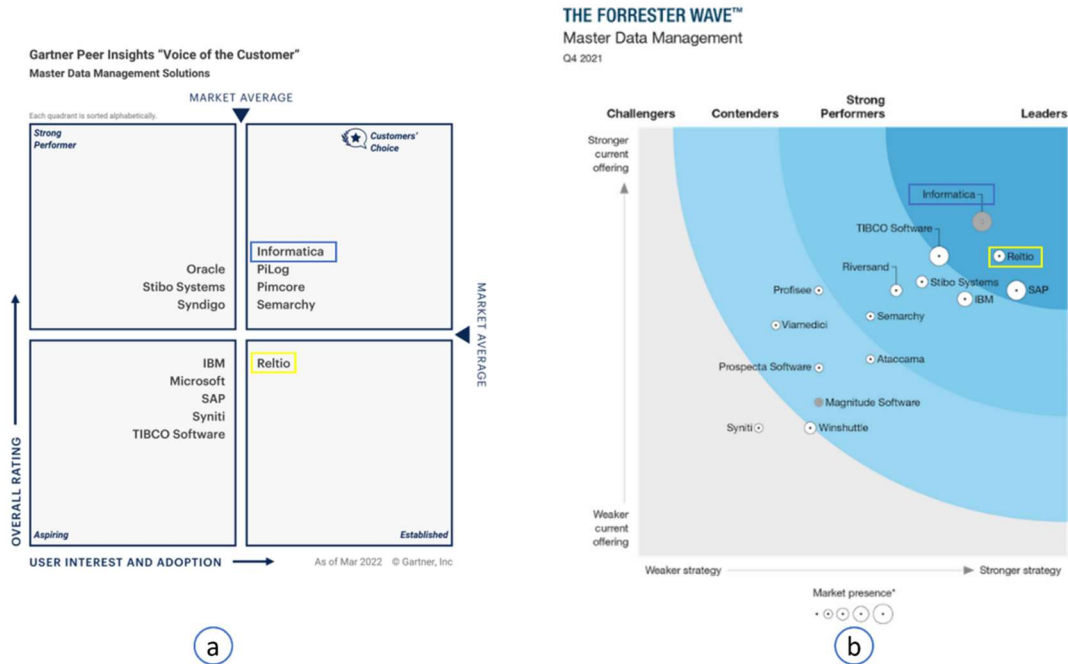


Figura 1. Cuadrante Mágico de Gartner – Soluciones para la Gestión de Datos Maestros - Año 2022 versus “The Forrester Wave Vendor Review” año 2021.

- **Cuadrante inferior derecho: Establecido**

“Proveedores cuyas calificaciones son inferiores a la calificación general promedio del mercado, pero tienen un interés y una adopción de los usuarios superiores al promedio. Los proveedores de este cuadrante recibieron calificaciones por debajo del promedio en cuanto a la experiencia general en relación con el mercado. Sin embargo, los proveedores en este cuadrante también superan el interés y la adopción de usuarios promedio del mercado, que incluye el volumen de reseñas, la cobertura del mercado de reseñas y la voluntad de recomendar.”

- **Cuadrante inferior izquierdo: Aspirantes**

“Proveedores que tienen calificaciones inferiores a la calificación general promedio del mercado y tienen un interés y una adopción del usuario inferiores al promedio. Los proveedores en este cuadrante están por debajo del promedio del mercado tanto en la calificación general como en el interés y la adopción de los usuarios. Sin embargo, los proveedores en este cuadrante pueden ser dignos de consideración para los usuarios, según los casos de uso específicos. Como todos los proveedores de este informe, sus productos se alinean con este mercado y han cumplido con los criterios mínimos para ser incluidos.”

Las siguientes características son las evaluadas por *Gartner Peer Insights* para clasificar las soluciones de gestión de datos maestros (MDM) (Gartner Peer Insights R&R, 2023).

- Admitir la identificación, vinculación y sincronización global de datos maestros en fuentes de datos heterogéneas a través de la reconciliación semántica de datos maestros.
- Crear y administrar un sistema central persistente de registro o índice de registro para datos maestros.
- Compatibilidad con los cuatro estilos de implementación del centro de MDM, según lo definido por Gartner.
- Habilitar la generación y entrega de una versión confiable de una o más áreas temáticas a todas las partes interesadas, en apoyo de diversas iniciativas de negocio.
- Apoyar los requisitos continuos de administración y gobierno de datos maestros a través de técnicas de acción correctiva y monitoreo basado en flujo de trabajo.
- Ser independientes del panorama de aplicaciones empresariales en el que residen.
- Las organizaciones de usuarios finales pueden implementarlas sin tener que recurrir a los profesionales necesarios.

La figura 1 (b) presenta el análisis de soluciones tecnológicas para la gestión de datos maestros realizada en el año 2021. La metodología de Forrester es descrita de forma general en (Forrester, 2023).

5. Diseño de la Arquitectura de Referencia

“Las arquitecturas de referencia para la gestión de datos maestros y de referencia ayuda a los administradores de proyectos, desarrolladores de software, arquitectos de datos y administradores de datos a comprender los datos en un contexto delimitado por el alcance de las empresas para gobernar los datos que son el eje de sus negocios.” (Cloutier, Muller, Verma, & Nilchiani, 2008).

5.1. Componentes de la Arquitectura de Referencia

La figura 2 muestra la arquitectura de referencia multidominio propuesta para la gestión de datos maestros para el Gobierno Colombiano. Esta se define desde cinco componentes principales: i) entidades fuentes y consumidoras, ii) Plataforma de interoperabilidad X-ROAD, iii) zonas de gestión de los datos maestros: llegada, staging, y verdad, y iv) componentes de apoyo a la gestión: gobierno de datos, linaje – trazabilidad, y v) Seguridad y protección de datos.

5.1.1. Entidades Fuentes y Consumidoras

Se definen dos tipos de entidades que interactúan con el Concentrador GDM.

Entidad Provedora. Es aquella que actúa como creadora y custodia de los datos maestros o de referencia que estarán relacionados en el Concentrador GDM, Para el alcance de este proyecto las entidades proveedoras son: Registraduría Nacional del Estado Civil, Confecámaras, e IGAC. Este tipo de entidad tiene toda la propiedad sobre sus datos en el dominio del cual ella es proveedora. El administrador de datos de la entidad (**rol definido en el marco de gobierno de la fase 2**) debe asegurar el cumplimiento de los requerimientos de gestión de datos maestros y de referencia definidos por el gobierno del Concentrador GDM) Además, debe aplicar principios de seguridad y protección de la información para cumplir con autenticación digital, si los datos de la entidad lo requieren.

Entidad Consumidora. Es aquella que consume datos maestros del Concentrador GDM. El acuerdo establecido con el Concentrador GDM determinará qué información puede consultar o acceder.

Los administradores de datos de ambos tipos de entidades deben crear, mantener, y administrar el gobierno de sus datos maestros y de referencia de acuerdo con los lineamientos, políticas, reglas y decretos establecidos por el gobierno de datos del Concentrador GDM.

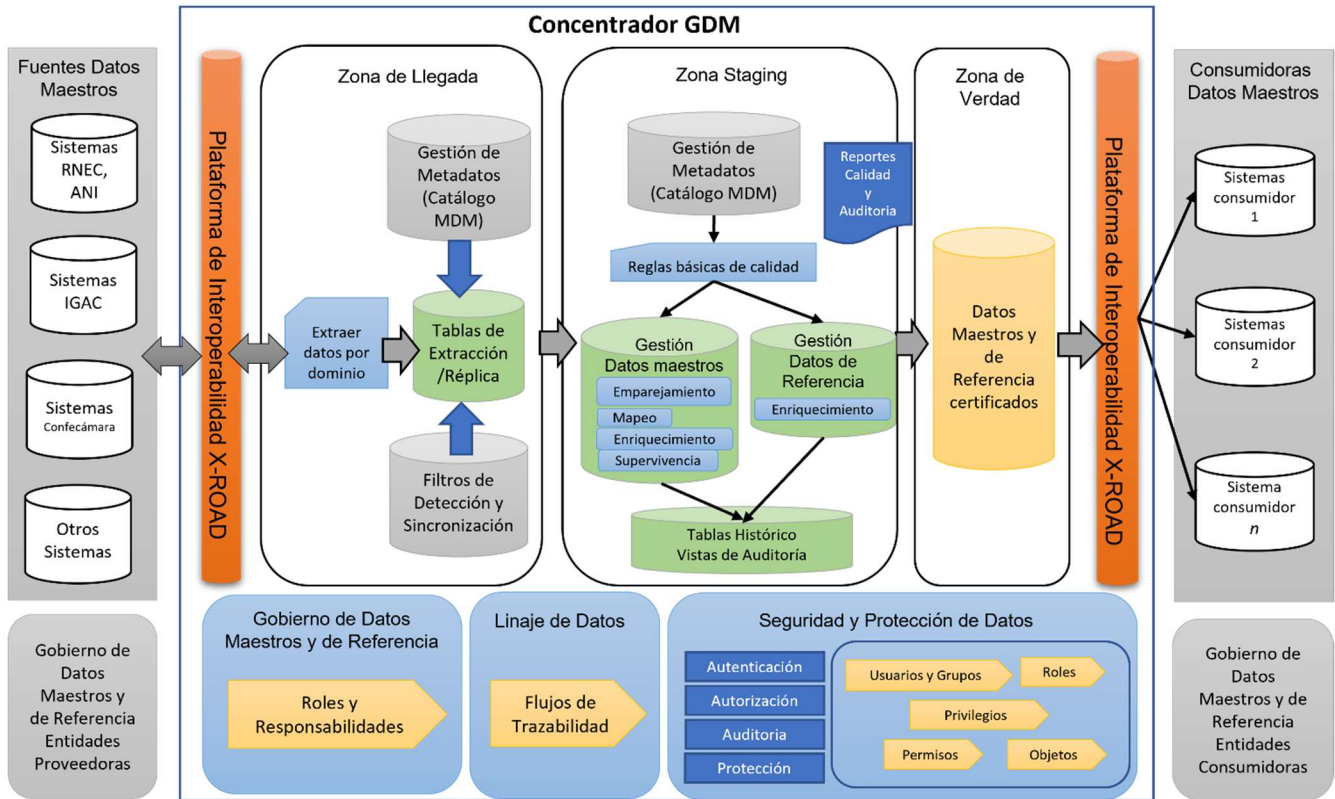


Figura 2. Arquitectura de Referencia para la gestión de datos maestros y de referencia en el Concentrador GDM. Fuente: Elaboración propia.

5.1.2. Plataforma de Interoperabilidad – X-ROAD

La figura 2 muestra gráficamente a la plataforma de interoperabilidad X-ROAD como la plataforma que asumirá el proceso de intercambio de datos para entidades proveedora y consumidoras. La tabla 3 presenta el conjunto de pasos que deben realizar las entidades y el Concentrador GDM para preparar el intercambio de datos que se establecerá para la extracción de datos de las entidades proveedoras y el consumo de datos por parte de las entidades consumidoras.

Tabla 3. Proceso de Intercambio de datos (una adaptación de (Minvivienda, 2021)).

Nro.	Actividad	Acción en la arquitectura de referencia
1	Alistamiento Plataforma XROAD	Las entidades proveedoras, para este caso: Registraduría, Confecámaras, IGAC, y el Concentrador GDM gestionan los requisitos y aprovisionamiento para integrarse con la plataforma X-ROAD
2	Alistamiento Plataforma Bus de servicios	Las entidades proveedoras/consumidoras y el Concentrador GDM deben de realizar el acuerdo de intercambio de información con base en las instrucciones definidas en el marco de interoperabilidad del Gobierno de Colombia (MinTIC - Marco de interoperabilidad, 2019). Este se centra en proveer mecanismos de intercambio que pueden ser realizados desde un servicio web, una API REST, o SOAP como protocolo de intercambio. Estos habilitan la extracción de datos en la entidad proveedora desde el Concentrador GDM y desde la entidad consumidora en el Concentrador GDM. Dependiendo del mecanismo que use la entidad o el Concentrador GDM se deben gestionar requisitos técnicos para el uso del bus de servicios.
3	Desarrollo de capa de intermediación en bus de servicios	Definir y desarrollar componente de intermediación en el bus de servicios
4	Ajustes a sistemas de información impactados	Ajustar sistemas de gestión de datos maestros u otros similares para consumir o exponer servicio web o usar REST o SOAP usando el bus de servicios.
5	Configuración de la plataforma X-ROAD y bus de servicios ambiente pruebas	Se debe habilitar el consumo o exposición de los servicios en la plataforma X-ROAD desde la arquitectura del Concentrador GDM de acuerdo con las políticas de uso de X-ROAD. Se debe incorporar la capa de intermediación (adaptador X-ROAD para bus de servicios) en el servicio expuesto en el bus de servicios.

Los roles del nivel táctico definidos en el marco de gobierno del Concentrador GDM deben asegurarse de que cada uno de los elementos definidos en cada dominio para la interoperabilidad se adapten a la arquitectura de referencia definida.

5.1.3. Zonas para la Gestión de Datos maestros y de referencia

La gestión de datos maestros y de referencia se divide en tres zonas para formalizar la división física que se debe asegurar para garantizar el [ciclo de vida de los datos maestros y de referencia](#), la gestión de linaje y consecuente trazabilidad. En cada zona, el Concentrador GDM de Datos garantiza la gestión de metadatos, y la auditoría de los datos maestros y de referencia. Estas zonas son:

- **Zona de llegada.** En esta zona se almacenan todos los conjuntos de datos o réplicas de tabla maestras o de referencia que se extraen desde cada dominio/entidad proveedora, mediante un proceso de ingesta de datos el cual apoyado por la plataforma X-ROAD, De esta forma, La zona de llegada habilita un esquema de datos o base de datos para las tablas de extracción una vez son validadas por reglas de detección y sincronización. En esta zona se realizan actividades relacionadas a:
 - La extracción de datos por dominio. Requerimiento de extracción de datos maestros o de referencia a la plataforma de interoperabilidad X-ROAD, por cada dominio. Esta función se activa

de acuerdo con una programación que se realiza en el Concentrador GDM para la extracción de datos en las entidades proveedoras.

- La aplicación de filtros de detección y sincronización es el conjunto de acciones requeridas para alistar los datos extraídos para la resolución, vinculación y consolidación de entidades de forma periódica en el Concentrador GDM.
 - El almacenamiento de la réplica de los registros de oro de cada dominio en tablas de extracción, así como los datos de referencia que son parte de dicho registro.
 - La actualización del [Catálogo de Datos](#), la cual ocurre cuando alguna estructura de datos es creada o alterada.
- **Zona de Staging.** En esta zona se crea el esquema de datos llamado **hubgdm**, el cual contiene los activos de datos que harán posible la gestión de los [datos maestros y de referencias](#) en el Concentrador GDM. En este esquema se almacenan los datos que fueron extraídos en la zona de llegada, una vez se transformen a partir de:
 - La aplicación de reglas básicas de calidad que guíen la normalización, y estandarización de datos. Cuando las entidades maestras y de referencia del Concentrador GDM, en cada dominio, son una réplica de las entidades maestras o referencia de las entidades proveedoras, no se requerirá la aplicación de este tipo de reglas. Esto debido a que los datos fuente deben estar gobernados bajo las políticas y reglas definidas por el gobierno de datos del Concentrador GDM (ver [ciclo de vida de los datos maestros y de referencia](#)). Cuando los datos de referencia son extraídos o cargados desde fuentes externas a las entidades proveedoras, se recomienda que se apliquen estas reglas de calidad.
 - La integración de los datos cuando un registro de oro definido en una entidad maestra o datos de referencia de cada dominio requiera de la aplicación de reglas de emparejamiento, mapeo, supervivencia y enriquecimiento principalmente.

En esta zona se independiza la gestión de los datos maestros y los datos de referencia. Las entidades maestras requieren de la ejecución de un conjunto de tareas relacionadas a frecuencia de sincronización, registro y uso de los datos. Por el contrario, los datos de referencia pueden extraerse desde las entidades, proveedoras, generarse al interior del Concentrador GDM, o ser capturados desde fuentes externas. De esta forma, la frecuencia de creación o actualización en el Concentrador GDM es eventual, así como su integración y sincronización

- La creación o actualización de datos de [auditoría](#)⁵, y [datos históricos](#), así como generación del reporte de calidad y auditoría, el cual debe estar disponible a través de una interfaz de consulta o reportería proporcionada por una solución tecnológica para la gestión de datos maestros.
 - La actualización del Catálogo de Datos, la cual ocurre cuando alguna estructura de datos es creada o alterada.
- **Zona de verdad.** En esta zona se define una base de datos para almacenar una copia de las entidades maestras y de referencia que fueron creadas y certificadas en la zona staging, para cada domino. Estas tablas maestras serán expuestas a las entidades consumidoras bajo reglas de presentación definidas y acordadas entre los administradores de datos del Concentrador GDM y de la entidad consumidora, bajo un acuerdo previamente establecido.

5.1.4. Componentes de apoyo a la gestión

El gobierno de datos maestros y de referencia, el linaje de datos – trazabilidad, y Seguridad y Protección de datos son componentes transversales al proceso, que rigurosamente deben llevar los datos hasta ser habilitados para su consumo. Cada uno de estos componentes se definen a continuación.

- **Gobierno de Datos Maestros y de Referencia.** Este componente corresponde al marco de gobierno para la gestión de datos maestros, el cual principalmente se centra en la gestión de roles y responsabilidades y otros componentes de aplicación que facilitan el gobierno de los diferentes activos de datos del Concentrador GDM (ver [Anexo2. Proceso Estrategia y Gobierno para la GDM Nivel Estratégico](#)). Este componente apoya la configuración y gestión de las actividades de control y supervisión que se deban realizar sobre los datos maestros y de referencia almacenados en el [esquema de datos hubgdm](#), además de otros activos de datos que posteriormente se implementen en la solución tecnológica.
- **Linaje de Datos – Trazabilidad.** El linaje de los datos cuenta la historia de los datos maestros y de referencia desde que se crean en las entidades proveedoras hasta que quedan estáticos al final de su transformación y uso en el Concentrador GDM. Es una actividad tipo servicio que permite definir la trazabilidad de los datos definiendo un linaje que se construye a partir de los metadatos y datos históricos y de auditoría que se registran cuando hay cambios en los datos y en las estructuras de los datos. Así, el linaje se convierte en un grafo en el cual los nodos son estados de transformación y los arcos son la unión con estados anteriores y posteriores. De esta forma, se pueden mantener flujos de trazabilidad para administrar los cambios que se hacen sobre las entidades maestras y de referencia, a través del tiempo.

⁵ La auditoría tiene como insumo la creación o actualización de datos en las tablas maestras o de referencia del Concentrador GDM.

- **Seguridad y Protección de Datos.** La gestión de los datos maestros y de referencia almacenados en el Concentrador GDM debe apoyarse en diferentes módulos de seguridad que incluyan los siguientes aspectos (Woody, 2013) (ver [Anexo 3. Proceso de Seguridad y Protección de Datos](#))
- **Autenticación.** Este módulo asegura que solo accedan a la información de cada dominio definido en el Concentrador GDM personas, grupos o aplicaciones correctos y aprobados. Es decir, para acceder al servidor de datos – Concentrador GDM, se deben activar estrictos controles de acceso de usuarios con mecanismos de autorización y autenticación. Es importante proveer al servidor de herramientas como *Active Directory*, Kerberos, LDAP, o claves a nivel de hardware para habilitar la autenticación multifactor. Estos aseguran que ningún usuario tenga acceso sin restricciones al servidor de la base de datos. La autenticación debe vincularse con la administración general de usuarios para garantizar que las credenciales de acceso se revoken cuando los usuarios cambien de tarea, abandonen la entidad o por cualquier otro motivo.

(MinTIC - SCD, 2020) define la autenticación digital como “el procedimiento que permite verificar los atributos digitales de una persona cuando adelanten trámites y servicios a través de medios digitales en las entidades, mediante el cual se puede afirmar que dicha persona es quien dice ser. Para el acceso a este servicio dichas entidades deben identificar y determinar el nivel de garantía a partir del grado de confianza requerido para sus procesos, siendo estos niveles de confianza clasificados como bajo, medio, alto y muy alto.”

- **Autorización.** El módulo de autorización de usuario y control de acceso permite implementar el acceso basado en roles para el Concentrador GDM. Es decir, permite que solo aquellos con privilegios específicos puedan tener acceso a ciertos datos. Los usuarios deben tener permisos para ver y trabajar con datos en la base de datos.

Desde el punto de vista de la carpeta ciudadana, (MinTIC - SCD, 2020) define la autorización como el “servicio que le permite a las personas naturales o jurídicas acceder y gestionar digitalmente de manera segura, confiable y actualizada al conjunto de sus datos, que tienen o custodian las entidades y solicitar actualización de esto.”

Este módulo de seguridad debe tener en cuenta los siguientes conceptos clave de control de acceso basado en roles:

- **Privilegios.** Un privilegio autoriza o deniega el acceso del usuario para realizar acciones en los activos de datos del Concentrador GDM.
- **Roles.** Un rol contiene uno o más privilegios del sistema donde cada privilegio define un derecho administrativo a un determinado objeto o tipo de objeto en el sistema. Al asignar un rol a un usuario, el usuario hereda las capacidades de los privilegios definidos en ese rol.
- **Usuarios y Grupos.** Los usuarios y grupos se utilizan para asignar roles desde un directorio activo.
- **Permisos.** Los permisos permiten asignar privilegios a usuarios o grupos para realizar ciertas acciones y realizar cambios en los objetos dentro del servidor de datos. Estos permisos afectan solo a los usuarios que inician sesión en el servidor de datos en lugar de los usuarios que inician

sesión desde una interfaz de la solución tecnológica, desde una página web, u otros servidores directamente.

- **Objetos.** Es una entidad sobre la cual se realizan acciones. Los objetos del Concentrador GDM son bases de datos o instancias, tablas, atributos, campos, índices, entre otros.
- **Auditoría.** El acceso a registros de auditoría registrada en el Concentrador GDM proporciona acceso a registros de auditoría y vigencia de los datos. Así se puede hacer seguimiento para ver quién ha solicitado acceso, o accedido, o creado y actualizado ciertos registros en el Concentrador GDM.
- **Protección.** Este módulo de seguridad debe proveer todos los elementos componentes requeridos para garantizar la protección de los datos maestros y de referencia almacenados en el Concentrador GDM. Dicha protección debe ser efectiva ante un posible evento de corrupción de los datos durante su ciclo de vida. Algunas de las funciones a tener en cuenta son:
 - **Cifrado del sistema de archivos.** La solución MDM (Master Data Management) debe proveer el sistema de cifrado para garantizar que el sistema de archivos sea protegido.
 - **Acceso a la red del servidor de la base de datos.** La solución MDM debe proveer la forma de configurar el control a las conexiones de autenticación al servidor de la base de datos. Normalmente esta configuración consiste en un conjunto de registros que definen el tipo de conexión, el rango de direcciones IP del cliente, un nombre de base de datos, un nombre de usuario, y el método de autenticación que se usará para las conexiones que coincidan con estos parámetros.
 - **Cifrado en tránsito y autenticación de servidor (SSL).** todos los datos, incluidas las contraseñas y los nombres de usuario, deben cifrarse en la red mediante SSL y los certificados garantizan que el usuario se comunica con la máquina host deseada.
 - **Prevención de ataques de inyección de SQL.** La solución tecnológica MDM debe asegurar esta función y que la utiliza para bloquear la corrupción o la cooptación de una base de datos, incluidas las relaciones no autorizadas, los comandos de utilidades, la tautología de SQL y el DML ilimitado.
 - **Cifrado de datos.** El cifrado de datos, en ciberseguridad, es el proceso que convierte la representación original de la información, conocida como texto plano, en una forma alternativa conocida como texto cifrado. El sistema de seguridad de la solución tecnológica debe garantizar que sólo las partes autorizadas pueden descifrar un texto cifrado para convertirlo en texto plano y acceder a la información original.”. El cifrado de datos a nivel de columna mantiene segura la información de la base de datos. (Kaspersky, 2023).
 - **Redacción de datos.** La solución tecnológica debe asegurar esta característica para proteger (o enmascarar) algunos elementos de datos de ciertos tipos de usuarios.

La tabla 4 muestra la descripción de posibles permisos que podrían declararse para los roles o grupos de roles definidos en el marco de gobierno para la gestión de datos maestros del Concentrador GDM. Además, la tabla 5 presenta una propuesta de cómo se asignarían estos permisos asociados a los diferentes grupos de roles.

Tabla 4. Descripción de permisos para el Concentrador GDM.

Permisos	Descripción
L: Leer	El usuario puede leer datos de los modelos del dominio al cual esta autorizado
C: Crear	El usuario puede crear estructuras de datos en el dominio al cual esta autorizado
A: Actualizar	El usuario puede actualizar estructuras de datos en el dominio al cual esta autorizado
E: Eliminar	El usuario puede eliminar estructuras de datos en el dominio al cual esta autorizado
H: Habilitar	Activar acceso a estructuras, datos, aplicaciones, o componentes del Concentrador GDM
D: Denegar	Denegar acceso a estructuras, datos, aplicaciones, o componentes del Concentrador GDM
V: Visualizar	El usuario puede visualizar tableros e informes
M: Monitorear	El usuario puede monitorear la gestión de los datos maestros y de referencia
Admin	Tiene todos los permisos para hacer operaciones CRUD en cualquier activo de datos definido en el Concentrador GDM

Tabla 5. Permisos para los grupos de roles definidos en el marco de gobierno del Concentrador GDM.

ROLES	Nivel Estratégico		Nivel Tactico				Nivel Operativo		
	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)
Componentes de Seguridad y Protección de Datos									
Permisos Concentrador GDM									
Control de acceso al servidor									
Acceso al servidor				Admin	L, V	V, M	L, C, A, E	M	V, M
Cifrado del sistema de archivos				Admin	L, V	V, M	L, C, A, E	M	V, M
Acceso a la red del servidor de la base de datos				Admin	L, V	V, M	L, C, A, E	M	V, M
Cifrado en tránsito y autenticación de servidor (SSL)				Admin	L, V	V, M	L, C, A, E	M	V, M
Prevención de ataques de inyección de SQL				Admin	L, V	V, M	L, C, A, E	M	V, M
Cifrado de datos				Admin	L, V	V, M	L, C, A, E	M	V, M
Redacción de datos				Admin	L, V	V, M	L, C, A, E	M	V, M
Control de acceso a los datos									
Tablas	V	V	V	Admin	L, V	V, M	L, C, A, E, D	M	V, M
Atributo/Columnas	V	V	V	Admin	L, V	V, M	L, C, A, E, D	M	V, M
Registros/Campos	V	V	V	Admin	L, V	V, M	L, C, A, E, D	M	V, M
Vistas	V	V	V	Admin	L, V	V, M	L, C, A, E, D	M	V, M
Barreras de Seguridad	V	V	V	Admin	L, V	V, M	L, C, A, E, D	M	V, M
Autenticación									
Usuarios				Admin					
Roles				Admin					
Perfilamiento de Contraseñas				Admin					

La tabla 6 presenta los permisos propuestos para diferentes tipos de usuarios de una plataforma de implementación del Concentrador GDM, con respecto a los componentes de seguridad y protección de datos.

Tabla 6. Permisos por Usuario del Concentrador GDM.

USUARIOS DEL CONCENTRADOR GDM	Tipos de Usuarios					
	Administrador	Read/Write	Operador de Datos	Seguridad y Protección	Proveedor datos	Consumidor datos
Componentes de Seguridad y Protección de Datos						
Permisos Concentrador GDM						
Control de acceso al servidor						
Acceso al servidor	Admin		V, M	V, M, H, D		
Cifrado del sistema de archivos	Admin		V, M	V, M, H, D		
Acceso a la red del servidor de la base de datos	Admin		V, M	V, M, H, D		
Cifrado en tránsito y autenticación de servidor (SSL)	Admin		V, M	V, M, H, D		
Prevención de ataques de inyección de SQL	Admin		V, M	V, M, H, D		
Cifrado de datos	Admin		V, M	V, M, H, D	M	
Redacción de datos	Admin		V, M	V, M, H, D	M	
Control de acceso a los datos						
Tablas	Admin	L, C, A, E	V, M	V, M, H, D	L, V, M	V
Atributo/Columnas	Admin	L, C, A, E	V, M	V, M, H, D	L, V, M	V
Registros/Campos	Admin	L, C, A, E	V, M	V, M, H, D	L, V, M	V
Vistas	Admin	L, C, A, E	V, M	V, M, H, D	L, V, M	V
Barreras de Seguridad	Admin	L, C, A, E	V, M	V, M, H, D	L, V, M	V
Autenticación						
Usuarios	Admin			V, M, H, D		
Roles	Admin			V, M, H, D		
Perfilamiento de Contraseñas	Admin			V, M, H, D		

- **Super usuario o usuario administrador.** El tipo de usuario tiene acceso completo a todas las aplicaciones, operaciones, estructuras, vistas, y atributos. Además, puede importar e implementar configuraciones de concentrador, ejecutar informes de análisis, establecer permisos de grupos de usuarios y ejecutar trabajos en el concentrador.
- **Usuario Read/Write.** Este tipo de usuario tiene acceso de lectura y escritura en todos los esquemas de bases de datos definidos en el Concentrador GDM. A este tipo de usuario se le otorgan permisos para realizar ediciones y actualizaciones de registros, editar estados de atributos, alterar datos resultado de reglas de supervivientes y otras definidas a nivel de la calidad de los datos. Además, podría agregar miembros a la base de datos, y otorga a otros tipos de usuario permisos para leer y escribir en todos o algunos esquemas definidos en el Concentrador GDM.
- **Usuario de Operación.** Este tipo de usuario tiene derechos de visualización y monitoreo de las tablas del Concentrador GDM desde la vista del front-end. Desde el back-end, este tipo de usuario es el encargado de crear y mantener métodos automatizados para orquestar la plataforma del Concentrador GDM. Además, monitorear la calidad de los datos y el estado de la infraestructura relacionada: las máquinas que operan las diferentes trabajos o procesos ETL, y canalizaciones (pipelines) para que la gestión de datos maestros y de referencia, así como con la interoperabilidad con las entidades proveedoras y consumidoras.
- **Usuario Seguridad.** Este tipo de usuario es el responsable de la seguridad de todos los activos de datos gestionados en el Concentrador GDM. Define los permisos y privilegios de acceso sobre los datos, estructuras, maquinas, y aplicaciones del Concentrador GDM. Esto implica la administración de usuarios, sus privilegios de acceso (administración de identidad y acceso), la protección de las cuentas en la nube contra el acceso no autorizado, el cifrado y la protección de los activos de datos y la administración de su cumplimiento.
- **Usuario Proveedor de Datos.** Este tipo de usuarios son externos al Concentrador GDM. Es el responsable del registro de oro y los datos de referencia que son extraídos del registro de origen. Una vez estos datos

son replicados en el Concentrador GDM, estos usuarios podrán leer, visualizar y monitorear el uso de sus datos desde la plataforma de implementación del Concentrador GDM.

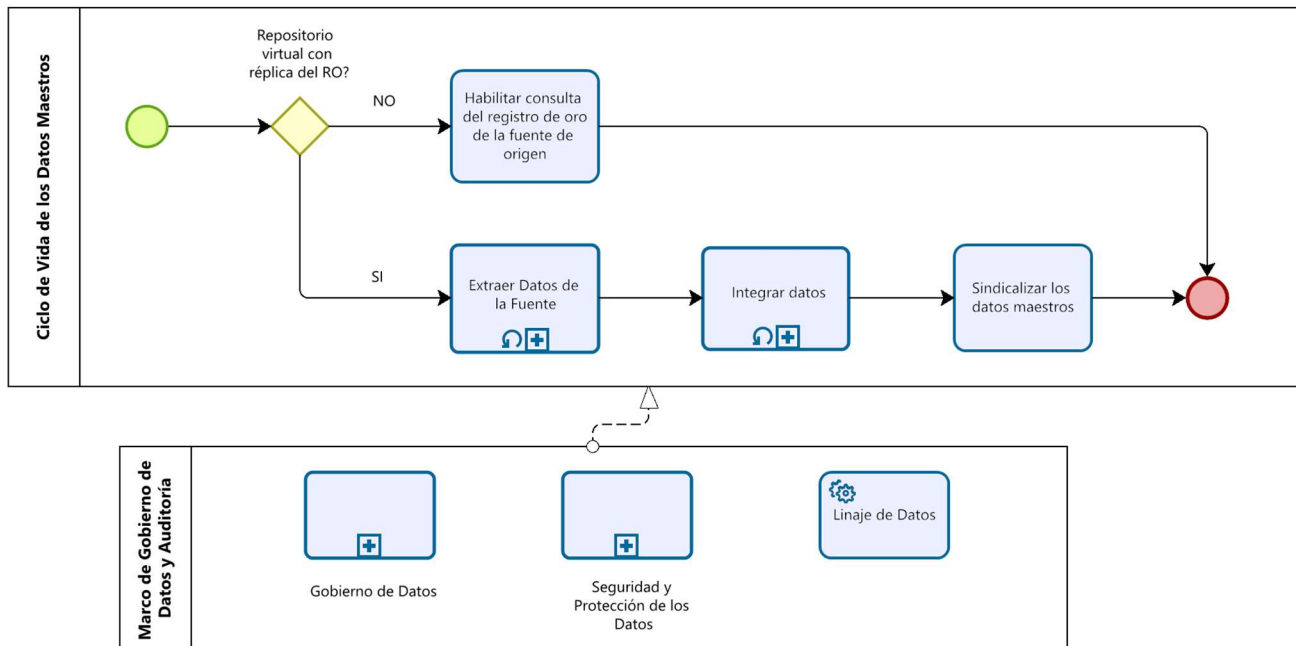
- **Consumidores de datos.** Este tipo de usuarios son externos al Concentrador GDM. Está autorizado a visualizar datos maestros almacenados en el Concentrador, de acuerdo con las reglas de presentación definidas en el acuerdo entre el Concentrador y la entidad consumidora.

5.2. Ciclo de Vida de los Datos Maestros y de Referencia

El ciclo de vida para la gestión de los datos maestros (CVGDM) le da vida a la arquitectura definida para el Concentrador GDM. Este ciclo es un mecanismo que permite monitorear y medir las diferentes acciones que se hacen sobre los datos maestros desde su creación hasta su puesta a disposición para las diferentes entidades consumidoras o usuarios autorizados. Los flujos de trabajo son diseñados bajo la notación BPMN (*Business Process Model and Notation*) (BPMN, 2023).

5.2.1. 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. En este caso, 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. La figura 3 muestra las actividades que forma parte del flujo del ciclo de vida de los datos maestros. Actividades asociadas al gobierno de datos, seguridad y protección de datos, y auditoría se definieron [sección 5.1.4](#).



Powered by
buzugi
Modeler

Figura 3. Proceso el ciclo de vida de la gestión de datos maestros. Elaboración propia.

Actividades del proceso

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:

a. Decisión: ¿repositorio virtual con réplica del Registro de Oro (RO)?

- Si no hay réplica: Se activa la siguiente tarea:

Habilitar consulta del registro de oro de la fuente de origen, En esta se define el acceso al 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.

- Si hay réplica. Se realizan las siguientes las tareas de los subprocesos de Extraer datos de la fuente, Integrar Datos, y Sindicalizar los Datos Maestros.

b. Subproceso: Extraer datos de la fuente

La figura 4 muestra el flujo de trabajo de este subproceso.

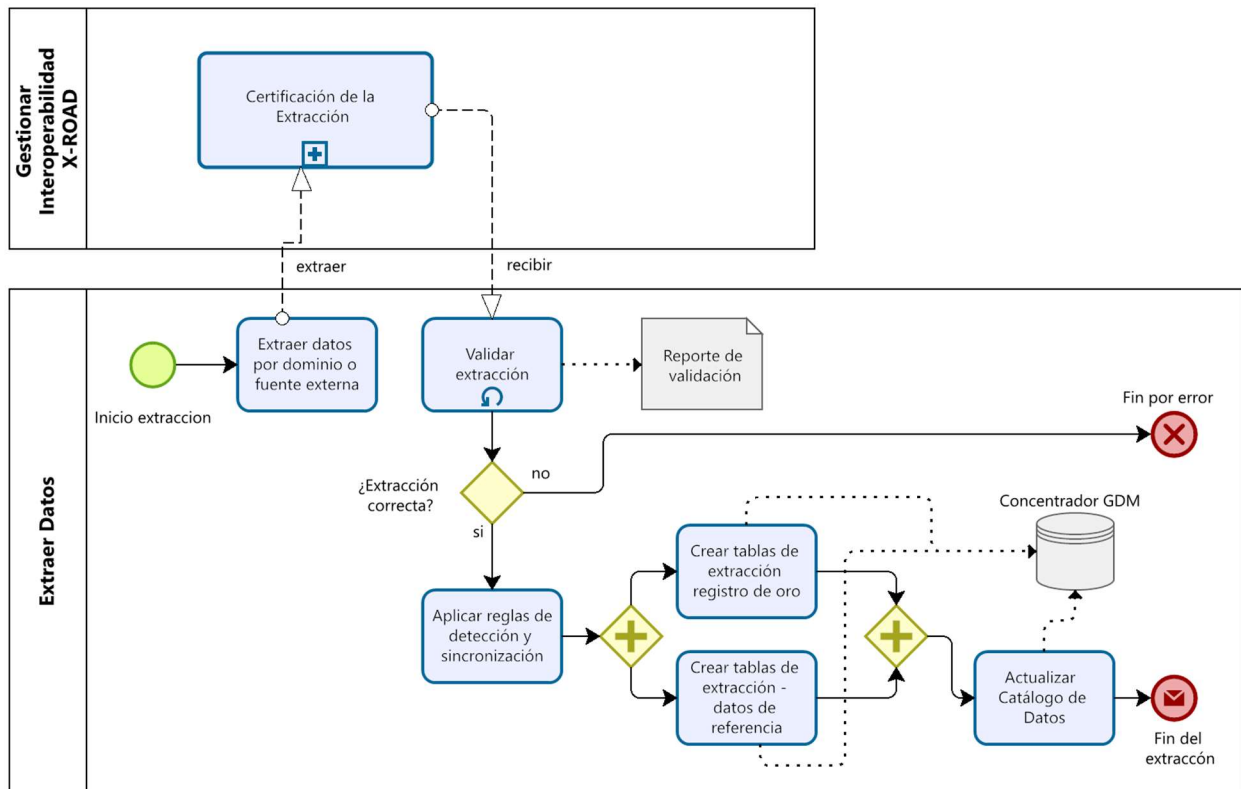


Figura 4. Subproceso: Extraer datos. Fuente: Elaboración propia.

Actividades del subproceso

- **Evento: Inicio Extracción.** El proceso inicia cuando el Concentrador GDM requiere extraer datos por cada dominio o desde una fuente externa. Esta soportado por un acuerdo que se hace entre cada entidad proveedora (fuente) y el Concentrador GDM a través de la plataforma X-ROAD. Los administradores de datos de cada una de las partes involucradas gestionan y supervisan el acuerdo. Dicho acuerdo debe establecer los datos a ser intercambiados y la periodicidad de extracción de los datos. Este proceso se ejecuta de forma iterativa hasta que termine la extracción de todos los registros de origen de cada uno de los dominios /entidad proveedora.
- **Extraer datos por dominio.** Desde la zona de llegada se activa una solicitud de extracción de datos por dominio la cual está agenda en un calendario de extracción, el cual es definido y supervisado por el

equipo de Operador Cloud del Concentrador GDM. Esta tarea envía un mensaje al proceso **Gestionar interoperabilidad X-ROAD** a través de un mensaje externo denominado “extraer”.

- **Mensaje externo: extraer.** Este mensaje le envía a la plataforma de interoperabilidad X-ROAD los siguientes metadatos con sus correspondientes valores:
 - a. Dominio de datos: <nombre del dominio>
 - b. Entidad proveedora: <nombre de la entidad donde se realiza la extracción>
 - c. Periodicidad de la extracción:<período> -- p.ej., diaria
 - d. Fecha de extracción programada: <dd/mm/aaa>
 - e. Usuario solicitante autorizado: <identificación usuario>
 - f. Lista de nombres de archivos a extraer (formato y nombre previamente estandarizados).
- **Certificación de la extracción.** Proceso externo realizado por la plataforma de interoperabilidad X-ROAD. Recibe el mensaje “Extraer” y ejecuta la extracción con base en los datos recibidos en el mensaje, los cuales usará como parámetros en alguno de los mecanismos de intercambio habilitados por la plataforma para realizar el intercambio de información. Al terminar la extracción, prepara y envía el mensaje “recibir” al Concentrador GDM con los datos extraídos.
- **Mensaje externo: recibir.** Este lleva cada uno de los archivos extraídos con los metadatos definidos en la tabla 7, con sus correspondientes valores.

Tabla 7. Características del registro de oro o datos de referencia de origen (ROO).

Fuente de verdad	Entidad: <entidad dueña o creadora de los datos fuente> Tabla/archivo: <nombre tabla certificada que contiene el registro de oro de origen> / Tabla/archivo: <nombre tabla certificada que contiene los datos de referencia del 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
Tipo de archivo	<maestro> o <referencia>
Metadatos del registro de oro de origen / datos de referencia de origen	Atributos del registro de origen
Metadatos auditoria por cada tupla del registro de oro de origen / tupla de datos de referencia	Fecha creación Fecha última modificación Evento sobre el registro (N:nuevo, M:modificado, Anulado) Nombre usuario certificador del evento
Tupla del registro de oro / tupla de datos de referencia	Conjunto de datos del registro de oro de la tabla maestra de origen o tabla de registro de referencia

- **Validar la extracción.** Recibe del mensaje “recibir” un lote de archivos que se extraen, con los datos maestros o de referencia, del registro de origen. Con base en este se procede a validar los datos en los siguientes niveles:

- Nivel de proceso: se evalúa que los datos esperados cumplan con los requisitos de calidad establecidos en el acuerdo a nivel del proceso.
- Nivel de archivo: se evalúa que archivos o tablas que se reciban sean los que se hayan convenido en el acuerdo establecido entre las partes a nivel de los archivos enviados vs recibidos.
- Nivel de registro: se evalúa que sean correctos los detalles de cada registro extraído, así como el número de registros.

Una vez termine la validación, genera un reporte de validación y envía un estado de la validación: Correcta o Incorrecta.

- **Decisión: ¿Extracción Correcta?**
 - Si no fue correcta: se genera un evento de fin con error que detiene el proceso del ciclo de vida de los datos que se han extraído.
 - Si fue correcta: se continúa con la tarea “Aplicar reglas de detección y sincronización”.
- **Aplicar reglas de detección y sincronización.** Esta tarea se requiere para alistar los datos en relación con la resolución, vinculación y consolidación en entidades de forma periódica (por ejemplo, todas las noches). Primero, se detectan datos extraídos de diferentes fuentes que van a contribuir al registro de oro de una misma entidad maestra definida en el Concentrador GDM. Si este es el caso, la sincronización se configura con períodos de tiempo exactos para que los gestores de datos maestros o de referencia de las entidades proveedoras no acumulen datos de diferentes períodos; así es posible 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.
- **Crear tablas de extracción registro de oro o datos de referencia.** Una vez se ha hecho la debida sincronización de entidades maestras y datos de referencia, se crean las tablas de extracción que almacenarán los datos maestros y los datos de referencia, por cada dominio, en el Concentrador GDM,
- **Actualizar Catálogo de datos.** Cuando se tienen listas las tablas de extracción, se procede a actualizar el catálogo de datos en el Concentrador GDM, según se requiera.
- **Evento: Fin de Extracción.** Este evento de fin del subproceso se define tipo mensaje; esto con el fin de dar continuidad al siguiente subproceso con los siguientes datos: el dominio, una lista con los siguientes metadatos: tipo de tabla (maestra o referencia), nombre de la tabla de extracción de extracción, período de extracción (día-mes-año).

c. Subproceso: Integrar datos

La integración de datos es un proceso que se realiza siempre y cuando una o varias fuentes certificadas del registro de origen aporten a la resolución de entidad del registro de oro en cada dominio o la integración de algún dato de referencia. Este proceso se realiza en el área *staging* para gestionar los [datos maestros y de referencia](#) del Concentrador GDM. La figura 5 muestra el subproceso de Integrar Datos.

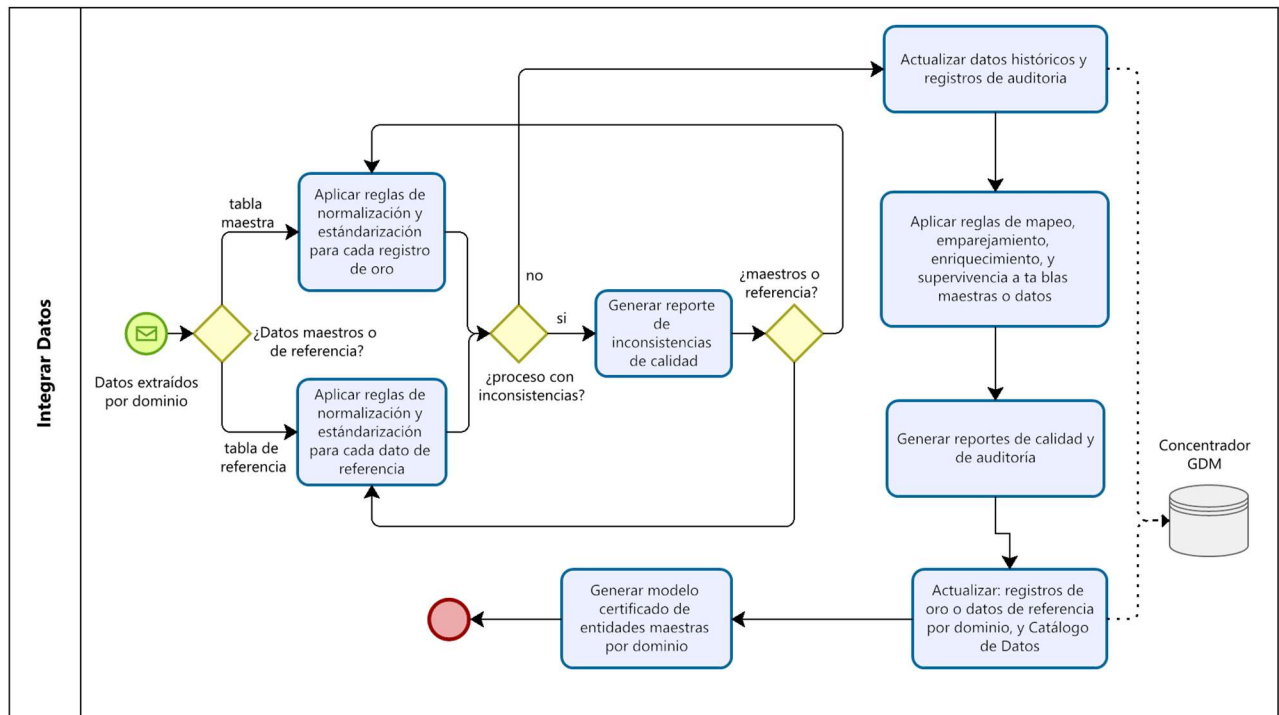


Figura 5. Subproceso: Integrar Datos. Fuente: Elaboración propia.

Actividades del subproceso

- **Evento Inicio: Datos Extraídos por dominio.** Este subproceso inicia una vez recibe el mensaje de fin del subproceso de extracción, con la información de fin del subproceso.
- **Decisión: ¿Datos maestros o de referencia?** Por cada ítem de la lista recibida, por dominio, la condición evalúa el tipo de tabla que recibe.
 - Si es una tabla maestra. El proceso se activa por el flujo de las tareas relacionadas a datos maestros para aplicar reglas asociadas para los registros de oro definidos de cada dominio.
 - Si es una tabla de referencia. El proceso se activa por el flujo de las tareas relacionadas a datos de referencia para aplicar reglas asociadas a los datos de referencia definidos por cada dominio o fuente externa.
- **Aplicar reglas de normalización y estandarización para cada registro de oro o para los datos de referencia.** De acuerdo con los datos del flujo recibido desde la condición, esta tarea procede de la siguiente forma: a cada registro de oro de una tabla maestra o registro de datos de referencia se aplican reglas de normalización y estandarización. Estas son reglas definidas en cada una de las dimensiones de calidad asociada a (ver anexo de dimensiones y reglas de calidad), las cuales hacen que el registro de oro o los datos de referencia que vienen de los registros de origen sean correctamente almacenados en el esquema hubgdm del Concentrador GDM. 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.
- **Decisión: ¿proceso con inconsistencias?** Una vez termine la aplicación de las reglas, es posible que algunas de ellas no se hayan podido ejecutar con éxito. Cuando esto se presenta, en la ejecución de cada regla se genera un mensaje de error. Esta situación activa un estado de inconsistencia con valor “Si”. Por el contrario, si todas las reglas se ejecutaron de forma correcta, el estado es “No”.
 - Si hay inconsistencias. Se continúa con la tarea “generar reporte de inconsistencias de calidad”.
 - No hay inconsistencias. Se continúa con la tarea “actualizar datos históricos registro de auditoría”.
- **Generar reporte de inconsistencias de calidad.** Con base en los resultados inconsistentes se prepara un reporte de calidad con las inconsistencias para el equipo táctico del concentrador GDM de tal forma que puedan tomar acciones correctivas y preventivas en conjunto con el Administrador de Datos de las entidades proveedoras, según sea el caso. Luego se activa la siguiente decisión.
- **Decisión: ¿maestros o referencia?** La decisión evalúa si el reporte es de:
 - maestros: entonces el flujo que lleva a la tarea “**aplicar reglas de normalización y estandarización por cada registro de oro**” para que se corrijan las inconsistencias reportadas.
 - referencia: entonces el flujo que lleva a la tarea “**aplicar reglas de normalización y estandarización por cada datos de referencia**” para que se corrijan las inconsistencias reportadas.
- **Actualizar datos históricos y registros de auditoría.** A medida que se van aplicando las reglas de calidad en cada uno de los registros de oro o en los datos maestros que lo requieran, se crean registros de datos históricos y se actualizan datos de auditoría en las tablas afectadas en el esquema **hubgdm**.
- **Aplicar reglas de mapeo, emparejamiento, enriquecimiento y supervivencia.** Según sea el caso requerido por cada domino, el grupo de reglas, [definido en el anexo Dimensiones y Reglas de Calidad](#), se aplican a los registros de oro de las entidades maestras o de los datos de referencia que requieran resolución de entidad. Estas tareas también **actualizan datos históricos y registros de auditoría**.
- **Generar reporte de calidad y auditoría.** Los reportes de calidad y auditoría se actualizan por los resultados arrojados por las tareas de “**Aplicar reglas de mapeo, emparejamiento, enriquecimiento y supervivencia**”, de tal forma que, el equipo táctico en conjunto con el administrador de datos de la fuente pueda tener control sobre este tipo de transformaciones.
- **Actualizar registros de oro y datos de referencia.** Una vez se actualice el reporte de calidad y auditoría, se genera una versión actualizada de los datos de las entidades maestras o de referencia.
- **Genera modelo certificado de entidades maestras por dominio.** Con base en la última versión generada de los registros de oro de cada entidad maestra en cada uno de los dominios, se procede a hacer una copia de esta versión al área de verdad. Las tablas o archivos de datos copia de los deben ser actualizados con metadatos que garantice la certificación.

d. Sindicalizar los datos maestros

Una vez se crean los datos maestros y de referencia en la zona de verdad o certificación para 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.

5.2.2. Ciclo de vida de los datos de referencia

Los datos de referencia pueden venir de fuentes externas⁶ o desde las entidades proveedoras para luego ser gestionados en el Concentrador GDM. El ciclo de vida de los datos de referencia se muestra en la figura 6. Actividades asociadas al gobierno de datos, seguridad y protección de datos, y auditoría se definieron [sección 5.1.4](#).

⁶ Una fuente externa es una organización que produce datos a nivel local, nacional o mundial que son de interés del gobierno Colombia para ser incluidos en el Concentrador GDM como datos de referencia.

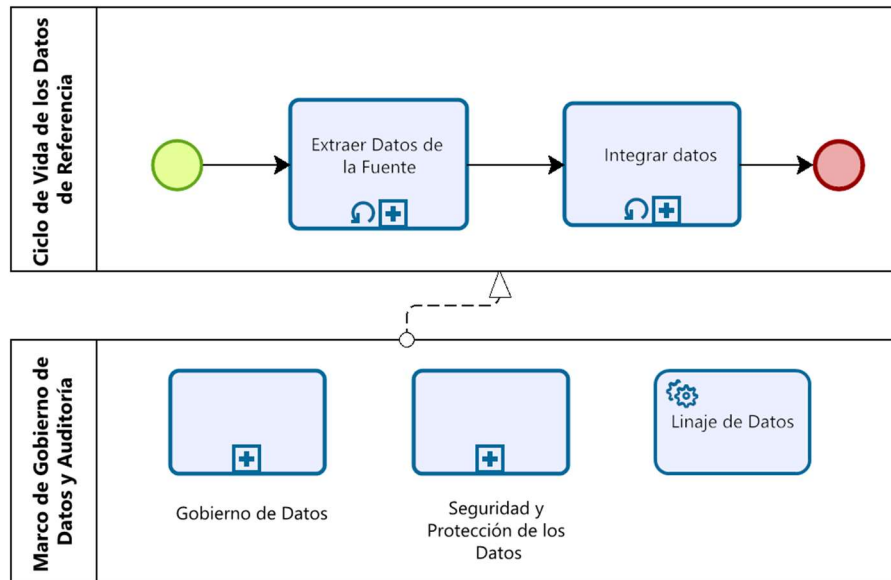


Figura 6. Ciclo de vida de los datos de referencia.

Actividades del proceso:

- **Extraer datos de la fuente.** Se reutiliza el subproceso Extraer Datos definidos en el ciclo de vida de los datos maestros.
- **Integrar datos.** Se reutiliza el subproceso Integrar Datos definidos en el ciclo de vida de los datos maestros.

El subproceso de Gobierno de Datos en el Marco de Gobierno de Datos y Auditoría se describen en el [Anexo 2. Proceso Estrategia y Gobierno para la GDM Nivel Estratégico.](#)

El subproceso de Seguridad y Protección de los Datos en el Marco de Gobierno de Datos y Auditoría se describen en el [Anexo 3. Proceso de Seguridad y Protección de los datos.](#)

El servicio de aplicación Linaje de Datos es una actividad tipo servicio que permite definir la trazabilidad de los datos definiendo un linaje que se construye a partir de los metadatos y datos históricos y de auditoría que se registran cuando hay cambios en los datos y en las estructuras de los datos.

Cualquier fuente externa que alimente 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.

5.3. Análisis de consideraciones especiales y recomendaciones

La arquitectura multidominio propuesta es evolutiva y flexible de tal forma que puede adaptarse fácilmente a decisiones de gobierno definidas entre el Concentrador GDM y las entidades proveedoras y consumidoras.

A medida que se va incrementando el número de entidades proveedoras y consumidoras que son parte del Concentrador GDM, es recomendable analizar la necesidad de establecer registros de oro más robustos para cada dominio. Esto implica identificar las entidades maestras y de referencia complementarias por cada dominio.

6. Modelo de Datos del Concentrador GDM

El modelo de datos diseñado para el Concentrador GDM está compuesto por tablas administrativas, tablas maestras, y tablas de referencia. Estas tablas permitirán mantener la gestión y el gobierno de estos datos en la arquitectura de referencia propuesta. La figura 9 muestra el modelo general el cual puede verse detalladamente en el archivo *hubgdm.sql* (anexo a este documento). Además, el modelo completo se describe completamente en el Catálogo de Datos,

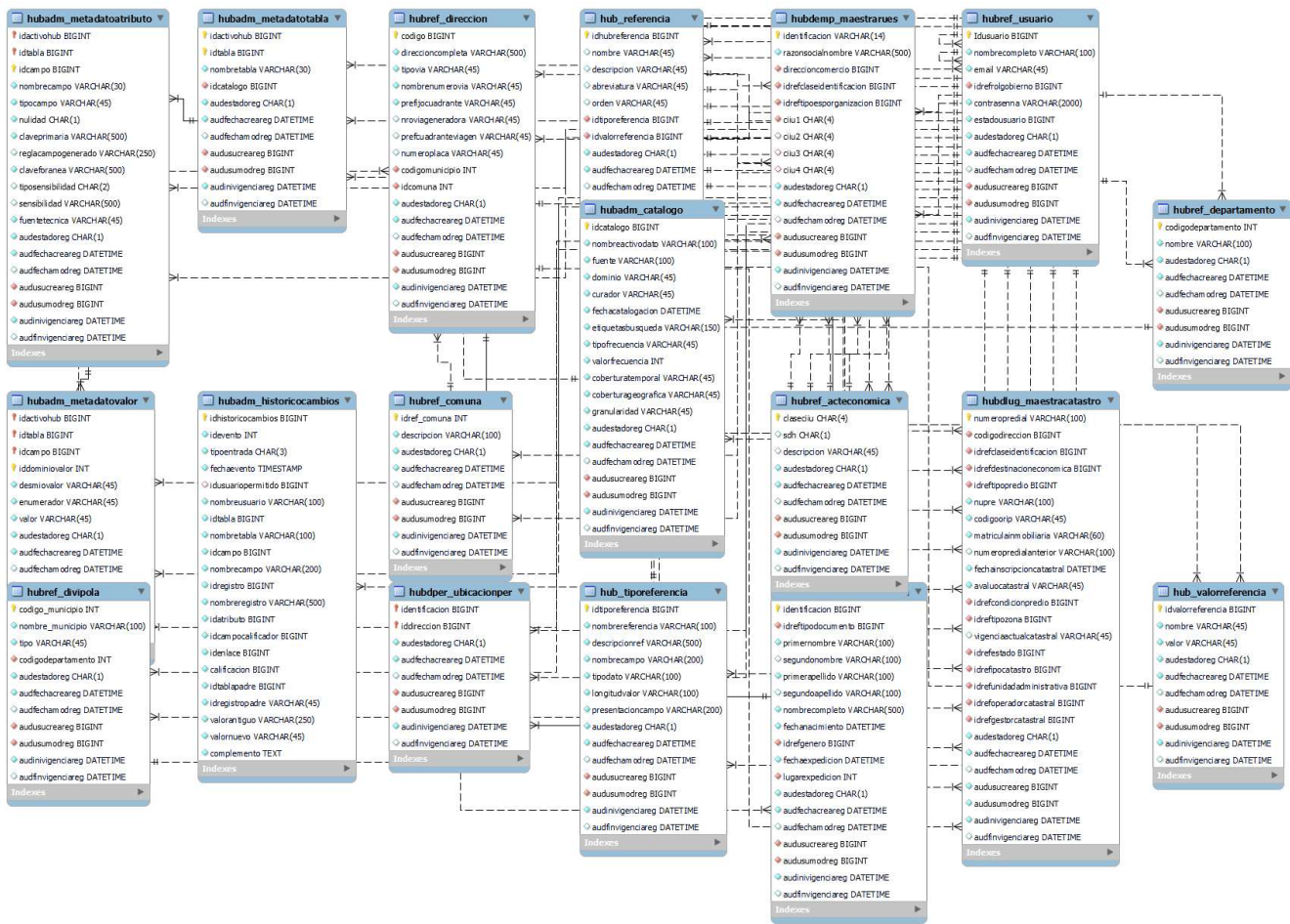


Figura 7. Modelo de Datos del Concentrador GDM. Fuente: Elaboración Propia.

Las reglas de calidad que se aplican a cada una de las entidades maestras, y de referencia se definen en el documento del diseño conceptual de la Fase 2. La tabla 6 muestra el estándar definido para el nombramiento de las tablas del modelo de datos.

Tabla 8. Tablas del modelo del Concentrador GDM

Prefijo	Descriptor	Tablas
Hubadm_<nombretabla>	Tablas para la administración del esquema/base de datos y de los datos almacenados en este.	<ul style="list-style-type: none"> - hubadm_historicocambios - hubadm_catalogo - hubadm_metadatotabla - hubadm_metadatoatributo - hubadm_metadatovalor
hubd<abreviación dominio>_<complemento>	Entidades maestras que almacenan datos maestros. Es la concatenación de dos palabras: “maestra” seguida de la palabra que define el nombre de la tabla maestra	<ul style="list-style-type: none"> - hubdper_maestrapersona - hubdemp_maestrarues - hubdlug_maestracatastro
Hubref_<nombretabla> o hub_<nombretabla>	Tablas que almacenan datos de referencia.	<ul style="list-style-type: none"> - hubref_acteconomica - hubref_direccion - hubref_divipola - hubref_departamento - hubref_comuna - hubref_usuario - hub_tiporeferencia - hub_valorreferencia - hub_referencia

6.1. Auditoria de Cambios

Cada una de las tablas que son parte del modelo de datos presentado en la figura 9 son definidas con atributos de auditoría, los cuales permiten monitorear los cambios que se hacen a los datos almacenados en las tablas. Los atributos de auditoría definidos en cada tabla se identifican por el prefijo **aud** antepuesto al nombre del atributo. Estos son:

- **audestadoreg**: estado del registro auditado: activo, inactivo
- **audfechacreareg**: fecha de creación del registro en la tabla
- **audfechamodreg**: fecha de modificación del registro en la tabla
- **audusucreareg**: usuario del Concentrador GDM que crea el registro en la tabla
- **audusumodreg**: usuario del Concentrador GDM que modifica el registro en la tabla

La vigencia de un registro de oro o de un dato de referencia tiene gran impacto en el uso de dicho registro en sistemas consumidores ya sean transaccionales y de analítica. Cuando un registro en el atributo **audfinvigencia** es diferente a nulo, es porque dicho registro deja de tener valor en el Concentrador GDM. Para controlar este tipo de estado se disponen de los siguientes atributos:

- **audinivigenciareg**: fecha de inicio de la vigencia del registro
- **audfinvigenciareg**: fecha fin de la vigencia del registro

La tabla **hubadm_historicocambios** no tiene los atributos de auditoría ya que es una tabla cuyos registros se crean al activarse un evento de cambio en los datos de una tabla o en su estructura. En otras palabras, los atributos de auditoría son la base de los registros almacenados en la tabla **hubadm_historicocambios**.

6.2. Tablas Administrativas

Las tablas administrativas definidas en el modelo de datos del Concentrador GDM tienen como objetivo habilitar una serie de estructuras para almacenar datos relacionados con los datos históricos y el catálogo de datos. El estándar de nombramiento de estas tablas es **hubadm_<nombratabla>**:

- Prefijo: **hubadm_**
 - o “**adm**”: abreviación que indica que la tabla es para apoyar la administración del Concentrador GDM.
- Complemento: **<nombratabla>** etiqueta definida por un sustantivo que indica el objetivo de almacenamiento de la tabla. La tabla **hubadm_catalogo** almacena los activos de datos definidos en el Concentrado GDM.

6.2.1. Modelo de datos históricos

Los datos históricos se crean cuando se realiza una operación de escritura (insertar, actualizar, eliminar) sobre alguna de las estructuras: tablas y atributos de las entidades maestras o de referencia o en los datos almacenados en el esquema de datos **huggdm**. La figura 10 presenta el modelo de datos históricos el cual está compuesto por la tabla **hubadm_historicocambios** y la tabla de referencia **hubref_usuario** que permite validar la existencia del usuario que hace el cambio.



Figura 8. Modelo de datos histórico de cambios.

La tabla 8 muestra el diccionario de datos de la tabla **hubadm_historicocambios** que almacenará el registro histórico de cambios.

Tabla 9. Diccionario de Datos de la tabla **hubadm_historicocambios**.

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
<u>Idhistoricocambios</u>	Valor único autoincremental	BIGINT	No	Si	No	No
Idevento	Evento que produce la creación del registro histórico: Insert, Update, Delete en cualquiera de las tablas que hacen parte del esquema hubgdm .	INT	No	No	No	No
Tipoentrada	El tipo de entrada define la acción de cambio realizada en el registro (ver posibles valores en la siguiente tabla).	CHAR(3)	No	No	No	No
Fechaevento	Fecha y hora en el que sucede la actualización de cualquiera de los campos que han sido alterados. En cada tabla del modelo, los atributos auditoría: audfechacreareg o audfechamodreg proporcionan la información base de este campo.	TIMESTAMP	No	No	No	No
Idusuariopermitido	Identificación del usuario permitido que ejecuta el comando u operación sobre la estructura o los datos. Clave foránea no obligatoria. En cada tabla del modelo, los atributos auditoría: audusucreareg o audusumodreg proporcionan la	BIGINT	No	No	hubref_usuario (idusuario)	No

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
	información base de este campo.					
Nombrequesuario	Nombre del usuario que realiza el cambio cuando se actualiza alguno de los campos rastreados	VARCHAR(100)	No		hubref_usuario (nombrecompleto)	No
Idtabla	Identificación interna de la tabla. Valor definido en el Catálogo de datos - Metadatos - Nivel Tabla o el que es asignado por el catálogo al interior de la base de datos.	BIGINT	No	No	No	No
Nombretabla	Nombre de la tabla en la que se ha cambiado un registro.	VARCHAR(100)	No	No	No	No
Idcampo	Identificación permanente del atributo. Valor definido en el Catálogo de datos - Metadatos - Nivel atributo	BIGINT	No	No	No	No
Nombrecampo	Nombre del campo del registro cambiado en la tabla origen	VARCHAR(200)	No	No	No	No
Idregistro	Identificación de registro permanente	BIGINT	No	No	No	No
Nombregistro	Contiene el valor del registro donde se genera el cambio	VARCHAR(500)	No	No	No	No
Idatributo	ID de atributo permanente	BIGINT	No	No	No	No
idcampocalificador	ID de campo calificador permanente	BIGINT	No	No	No	No
Idenlace	ID de enlace de datos calificados	BIGINT	No	No	No	No
Calificacion	Valor de atributos numéricos	BIGINT	No	No	No	No
Idtablapadre	ID de tabla del registro principal	BIGINT	No	No	No	No
Idregistropadre	ID de tabla del registro principal	VARCHAR(45)	No	No	No	No
Valorantiguo	Es valor del campo antes de que se cambie. Cuando se agrega un registro, no hay un valor anterior; en su lugar, se registra <Agregado>	VARCHAR(250)	No	No	No	No

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foránea	Sensibilidad
Valornuevo	Es el valor del campo después de que se cambie. Si se elimina un registro, ya no hay un valor; en su lugar, se registra <Eliminado>	VARCHAR(250)	No	No	No	No
Complemento	Si un valor registrado es demasiado largo y no se ajusta al límite de los campos valorantiguo y valornuevo , la parte que no se ajusta se registra aquí.	TEXT	No	No	No	No

El atributo “**tipodeentrada**” registra la operación de cambio la cual está definida por un conjunto de valores asociados al tipo de cambio ya sea en el valor de un campo, a nivel de atributo o a nivel de tabla.

- e. Tipos de entrada (**I**: insertar, **M**: modificar, y **E**: eliminar). Este tipo de cambio son cambios de nivel de registro estándar, es decir la tupla afectada por la operación de escritura. Los campos de la tabla **hub_historicocambios** están relacionados con los cambios que ocurren en las tuplas o registros de las tablas que hacen parte de las entidades maestras y las entidades de datos de referencia.

Los campos que representan este tipo de entrada **I**-Insertar, **M**-Modificar y **E**-Eliminar, en el caso de tuplas son:

- **idtabla** rastrea el ID de instancia de tupla donde se realiza el cambio
- **idregistro** contiene el Id del registro de la tupla que es cambiada (extraída de la tabla origen).

- f. Tipos de entrada (**IA**, **MA** y **EA**): cambios de atributos (**A**) de nivel de registro. Estos cambios son similares a los cambios de nivel de registro estándar (**I**, **M** y **E**). La principal diferencia es:

- **idcampo** contiene el ID permanente del campo donde se realiza el cambio. Esto incluye posibles jerarquías de campos en la misma tabla.
- **idatributo** contiene el ID permanente del atributo.

Si el atributo es numérico, la calificación o valor alterado se almacena en la columna **calificacion** y si no es numérico se agrega a la columna **nombrecampo**. Por ejemplo, un atributo ‘estado’ podría tener varias calificaciones: Estado [Activo], Estado [Noactivo]

- Tipos de entrada (IC, MC y EC): cambios en el calificador de nivel de registro. Estos cambios también son similares a los cambios de nivel de registro estándar (I, M y B).
 - **idcampo** contiene el campo de aplicación con identificación permanente.
 - **idcampocalificador** contiene el campo calificador con identificación permanente.
 - **idenlace** contiene el vínculo identificador.
- Tipos de entrada (AA, MA, EA): cambios en la definición de atributos. Estos cambios en las definiciones de atributos podrían llegar ser para alguno de los siguientes metadatos: tipo, nombre, alias, definición, etc.
 - **idatributo** contiene el ID permanente del atributo.
 - **valorantiguo** contiene los metadatos del atributo antes de hacer el cambio.
 - **valornuevo** contiene los metadatos del atributo después de hacer el cambio.
- Tipos de entrada (AV) - Cambios de vínculo de atributo. Se utiliza cuando cambia el estado de un vínculo de atributo. **valorantiguo** y **valornuevo** usan 0 para indicar Desvinculado y 1 para indicar Vinculado.
- Tipos de entrada (IT, MT y ET): cambios en la definición de la tabla. Para estos cambios en la descripción de la tabla, los campos que se activan son **idregistro** y las propiedades del campo tienen **idcampo** y **nombrecampo**.
 - **valorantiguo** contiene los metadatos de la tabla antes de hacer el cambio
 - **valornuevo** contiene los metadatos de la tabla después de hacer el cambio.

6.2.2. Modelo del Catálogo de Datos

La figura 11 muestra el modelo de las tablas administrativas que permiten mantener un catálogo de datos. Estas facilitan la gestión de los metadatos del modelo de Concentrador GDM.

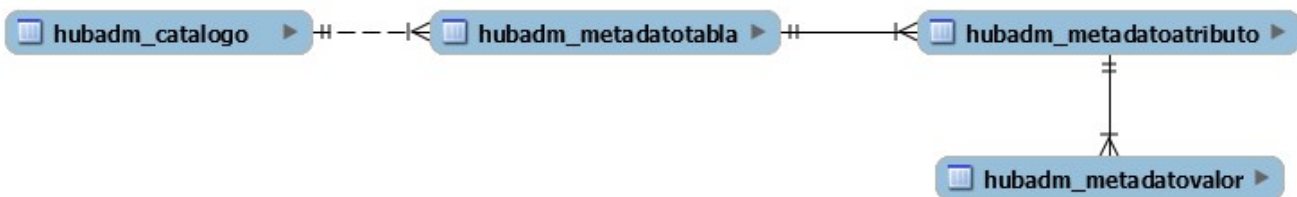


Figura 9. Modelo de datos de las tablas administrativas que son parte del Catálogo de Datos. Fuente: Elaboración Propia.

El diccionario de este modelo se define en el capítulo 6 – Gestión del Catálogo de Datos, en el documento de la Fase 2 – Diseño Conceptual.

6.3. Modelo de Datos – Tablas Maestras y de Referencia

La figura 12 muestra el modelo de las tablas maestras y de referencia que son parte del esquema hubgdm del Concentrador GDM. Las tablas maestras almacenan el registro de oro de cada dominio y las tablas de referencia almacenan los datos de referencia que complementan a las entidades maestras.

El estándar de nombramiento de tablas maestras es **hubd<abreviación dominio>_<complemento>** y se debe interpretar de la siguiente forma.

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

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 GDM
 - 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.

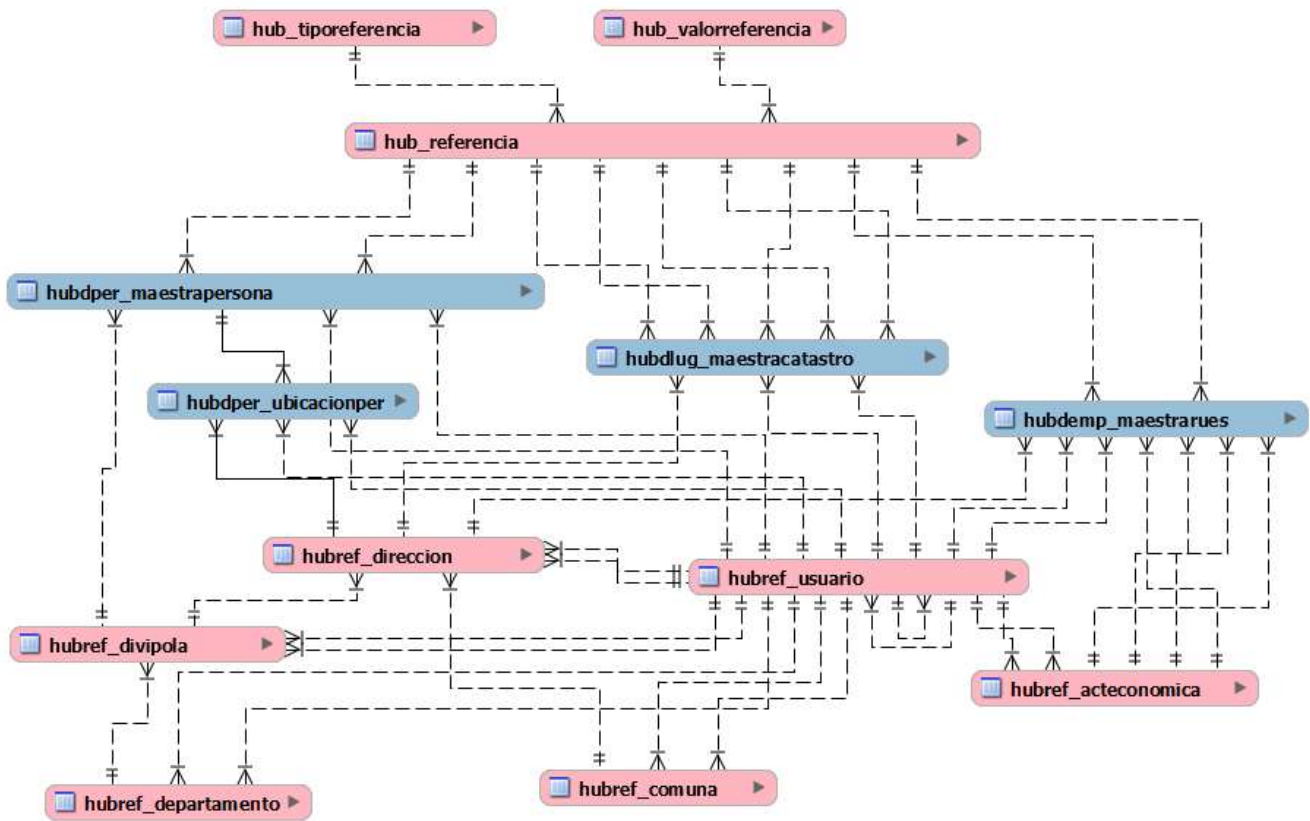


Figura 10. Modelo de Datos – Tablas Maestras y de Referencia. Fuente: Elaboración Propia.

6.1. Tablas Maestras

Las tablas maestras definidas en el modelo de datos del Concentrador GDM son entidades maestras que tienen el objetivo de habilitar una serie de estructuras para almacenar datos relacionados con los datos maestros proporcionados por las fuentes de datos. Como se presentó en el documento de la Fase 2 – Diseño Conceptual, capítulo 4 – sección Modelos de Datos, las entidades fuentes deben proporcionar un registro de oro que pueda ser replicado de forma exacta en el Concentrador GDM.

Cuando el registro de oro de referencia (ROR) del Concentrador GDM es construido por atributos proporcionados por una única entidad fuente certificada, el proceso de resolución de entidad maestra no es necesario ya que dicha entidad de origen provee la única fuente de verdad. Este es el caso de las entidades maestras de los dominios de Personas, Empresas, y Lugares.

Cuando el registro de oro de referencia del Concentrador GDM es construido por atributos que proveen diferentes entidades fuentes de verdad, es necesario activar el proceso de resolución de entidad en la tarea de integración de los datos definido en el ciclo de vida de los datos.

6.1.1. Modelo Maestra de Personas

La figura 13 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).

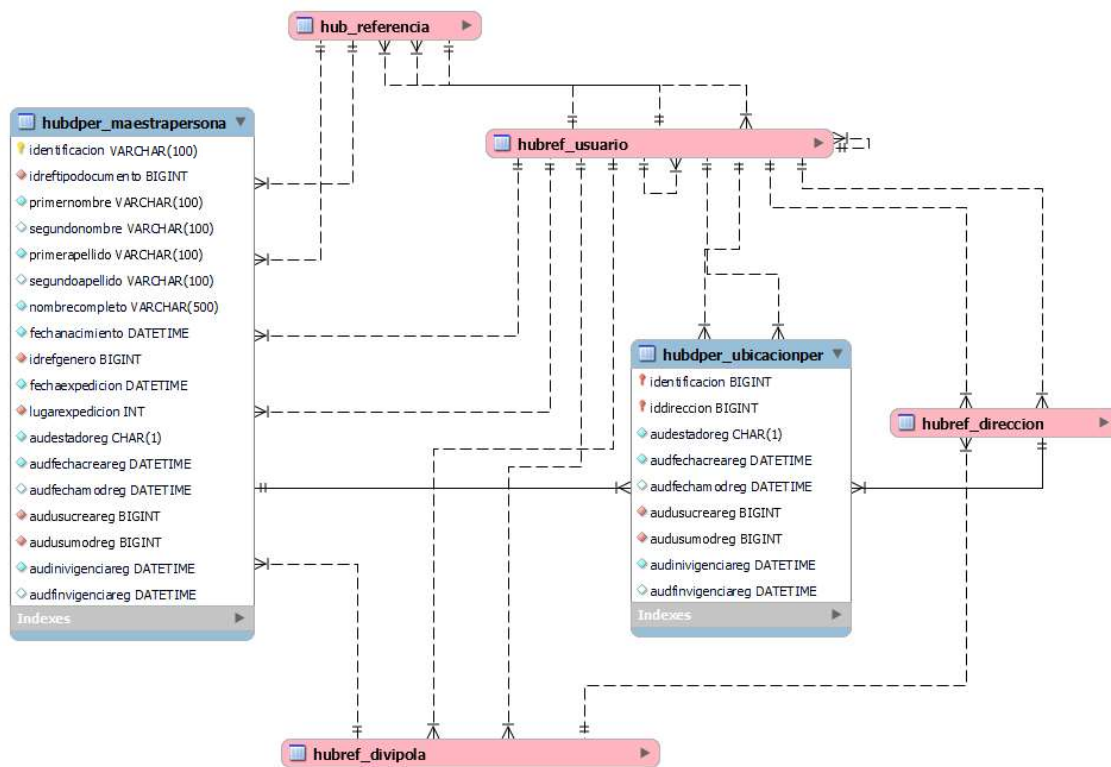


Figura 11. Modelo de datos de la tabla maestra de personas hubdper_maestrapersona – dominio de Personas.

La tabla **hubdper_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 **hubdper_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 8 presenta el diccionario de datos de la tabla maestra **hubdper_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 10. Diccionario de Datos Maestra de Persona – hubdper_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	VARCHAR(100)	No	Si	No	PII
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

Nombre Columna	Descripción Columna	Tipo de Dato	Es Nulo	Es Clave Primaria	Es Clave Foranea	Sensibilidad
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
Lugarexpedicion	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 9 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 11. 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	VARCHAR(100)	No	Si	hubdper_maestrapersona (identificación)	PII
Iddireccion	Tipos de documento definidos por RNEC	SMALLINT	No	No	Hubref_direccion (iddireccion)	No

6.1.2. Modelo Maestra de Empresa

La figura 14 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.

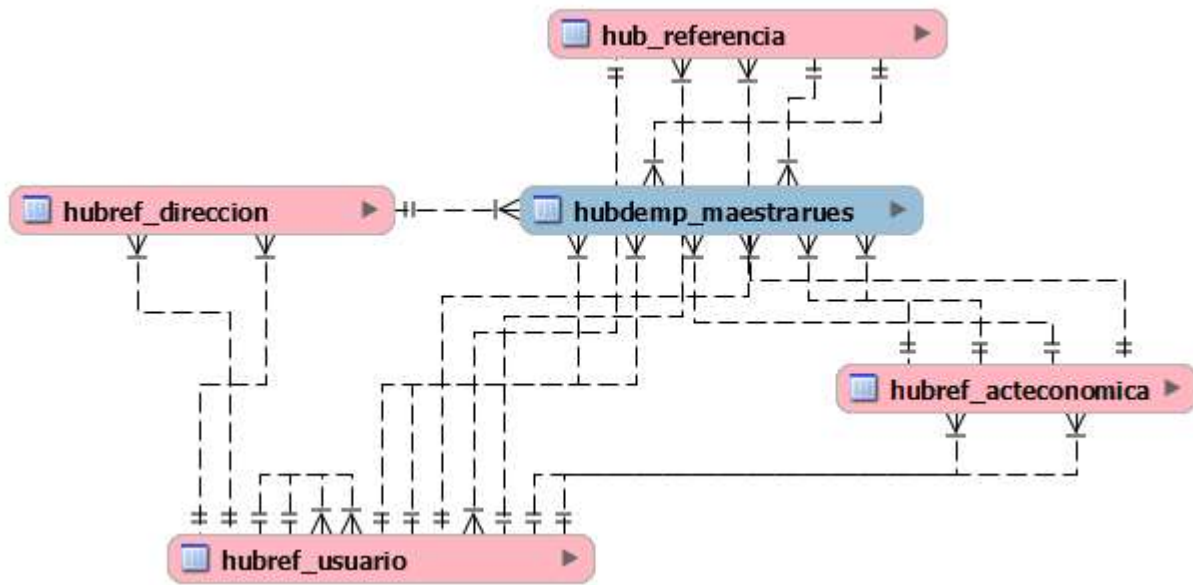


Figura 12. Modelo de datos de la entidad maestra de RUES en el ROR – dominio de Empresas.

La tabla 10 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(100)	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					

6.1.3. Modelo Maestra de Catastro

La figura 15 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.

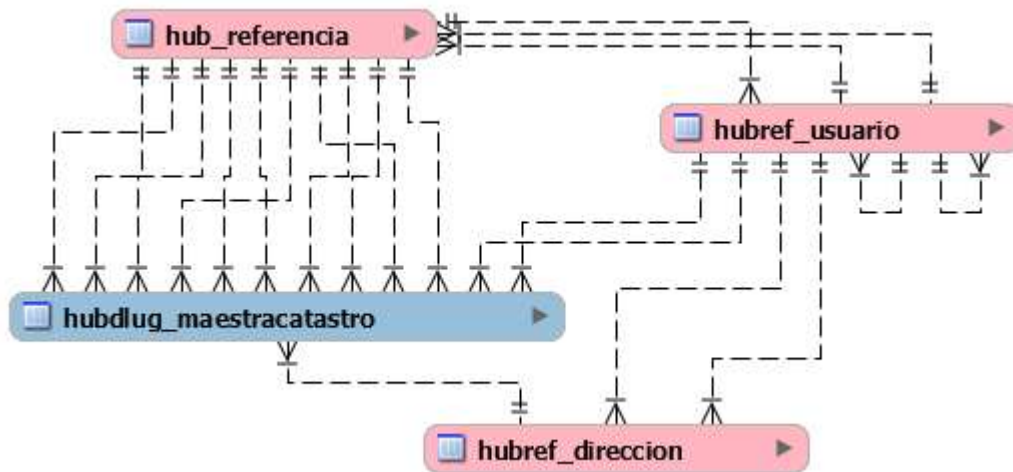


Figura 13. Modelo de Datos de la maestra de Catastro hubdlug_maestracatastro – dominio Lugares.

La tabla 11 presenta el diccionario de datos de la tabla maestra **hubdlug_maestracatastro**.

Tabla 13. 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

6.4. Tablas de Referencia

Las tablas de referencia definidas en el modelo de datos del Concentrador GDM son entidades de referencia que tienen el objetivo de habilitar una serie de estructuras para almacenar datos externos proporcionados por las entidades fuentes de datos o por otras instituciones que provean datos de referencia. Las tablas de referencia complementan los registros de oro definidos en las tablas maestras como lo hemos visto en la sección anterior.

6.1.4. Tabla de referencia – hubref_direccion

La figura 16 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**.

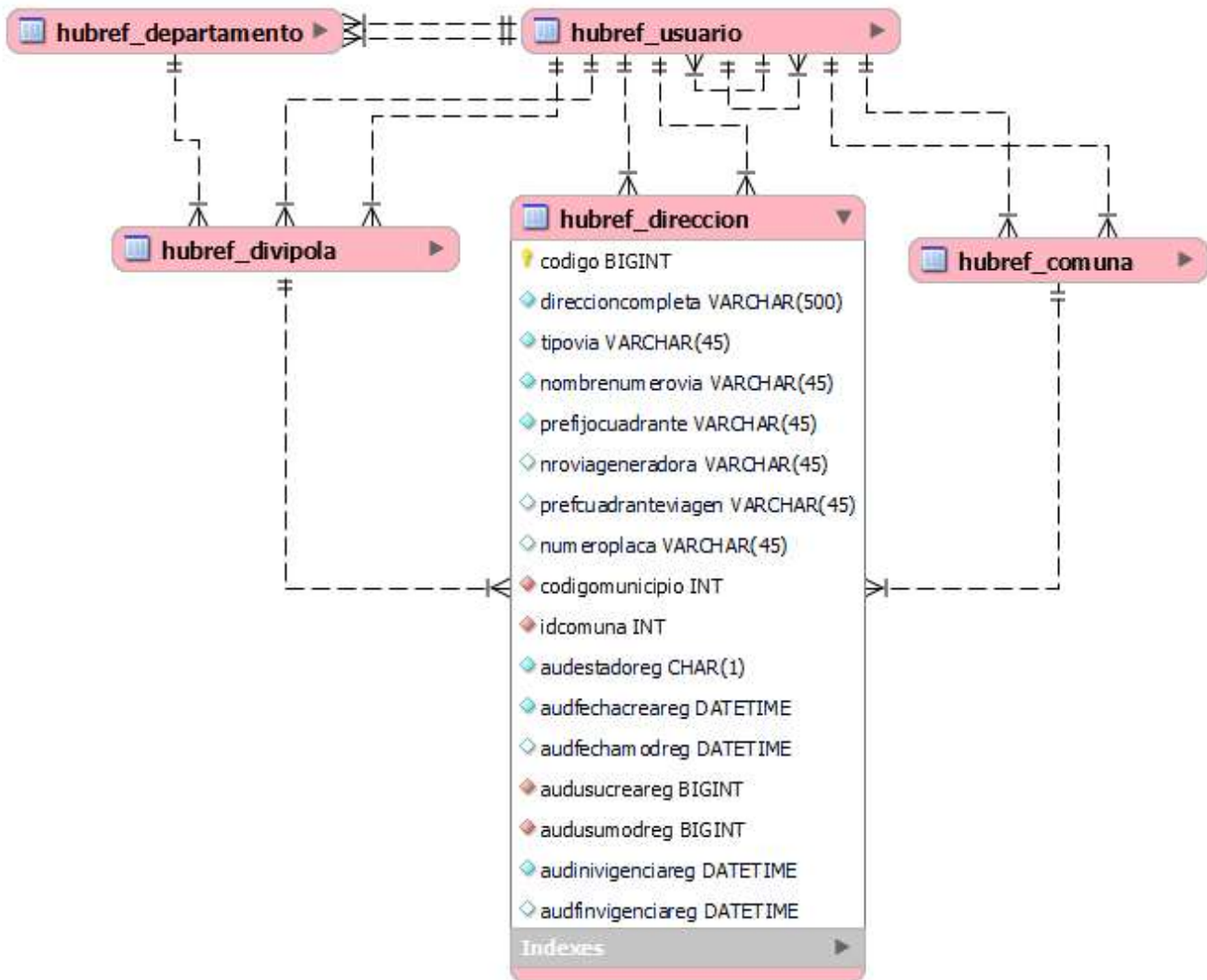


Figura 14. Modelo de Referencia de Dirección – hubref_direccion.

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

Tabla 14. 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 13, 14, y 15 presentan el diccionario de datos de la tablas de referencia **hubref_divipola** y **hubref_departamento**, y **hubref_comuna**.

Tabla 15. 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 16. 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 17. 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

6.1.5. Tabla de referencia – hubref_usuario

La figura 17 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.

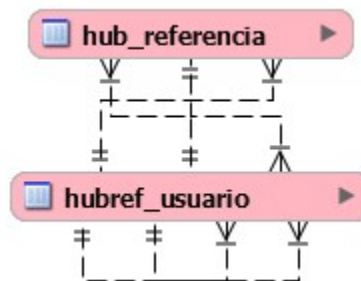


Figura 15. Modelo de Referencia de Usuario – Tabla hubref_usuario.

La tabla 16 presenta el diccionario de datos de la tabla de referencia **hubref_usuario**.

Tabla 18. 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

6.1.6. Tabla de referencia – hub_referencia

La figura 18 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.

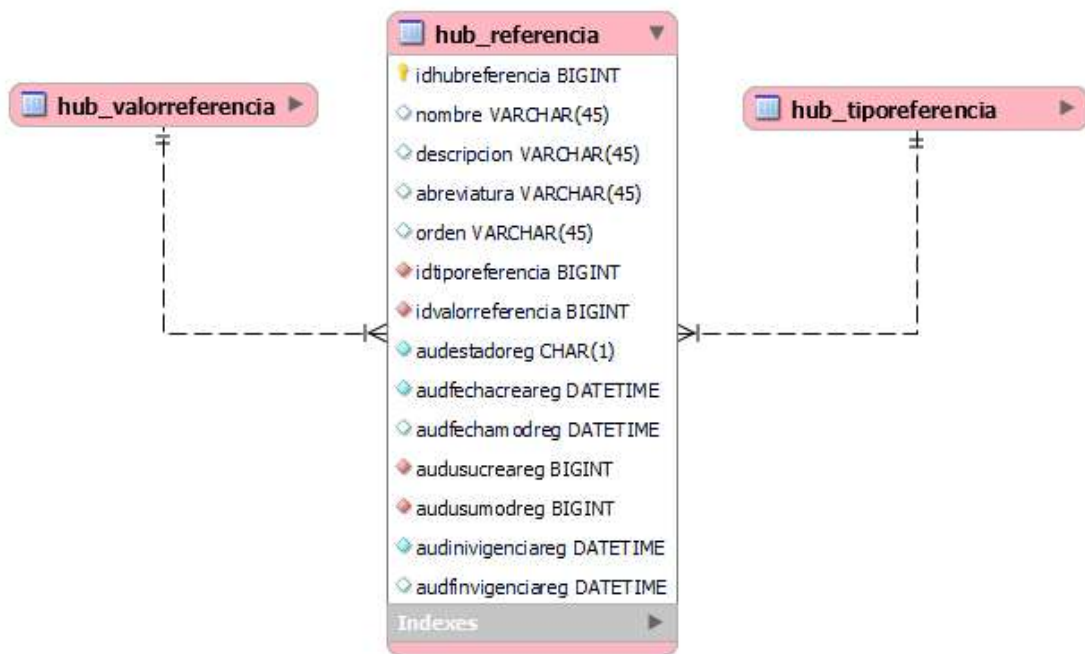


Figura 16. Modelo de Referencia – Tabla hub_referencia.

La tabla 17 presenta el diccionario de datos de la tabla de referencia **hub_referencia**.

Tabla 19. 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 18 presenta el diccionario de datos de la tabla de referencia **hub_tiporeferencia**.

Tabla 20. 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 19 presenta el diccionario de datos de la tabla de referencia **hub_valorreferencia**.

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
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

7. Análisis de Soluciones para la Gestión de Datos Maestros

La tabla 20 presenta un primer análisis de diez de las soluciones tecnológicas para la Gestión de Datos Maestros seleccionadas como base en la revisión y calificación que realiza (Gartner Peer Insights R&R, 2023). Este análisis se complementa con una revisión de la capacidad mínima asociada a una arquitectura multidominio y si la solución es 100% o no en su base tecnológica un MDM.

Tabla 22. Características de Mercado de la Soluciones MDM en Nube (Las mejores calificadas por Gartner Peer Insights).

Vendedor	Solución MDM	Ranking (5 estrellas)	Partners en Colombia Identificados	Cientes en Colombia Identificados	Centrado en MDM
Syndigo	Riversand MDM Center	71%	No identificado	No identificado	100%
PILog	PiLaog MDM	68%	No identificado	No identificado	80%
IBM	IBM MDM Cloud	67%	Si	Si	100%
Pimcore	Pimcore	63%	No identificado	No identificado	33%
Semarchy	Semarchy xDM	61%	Keyrus: https://keyrus.com/latam/es/home	No identificado	100%
TIBCO Software Inc.	TIBCO EBX	56%	it-Nova: https://it-nova.co/tibco-gestion-de-datos/	No identificado	100%
Profisee	Profisee Platform	55%	No identificado	No identificado	100%
Informatica	Intelligent Master Data Management Platform	52%	Keyrus: https://keyrus.com/latam/es/home	Si	100%
Reltio	Reltio Connected Data Platform	48%	No identificado	No identificado	100%

Los vendedores Informática y Semarchy, clasificados en el cuadrante “Elección del cliente” y resaltadas en color verde, no tienen la mayor calificación. Sin embargo, se ubican en el cuadrante de “Elección de los clientes” por tener una solución fortalecida desde características como la arquitectura técnica, gobernanza, calidad de los datos, entre otras. En la figura 19 se muestra la arquitectura de referencia MDM de la solución Informática.

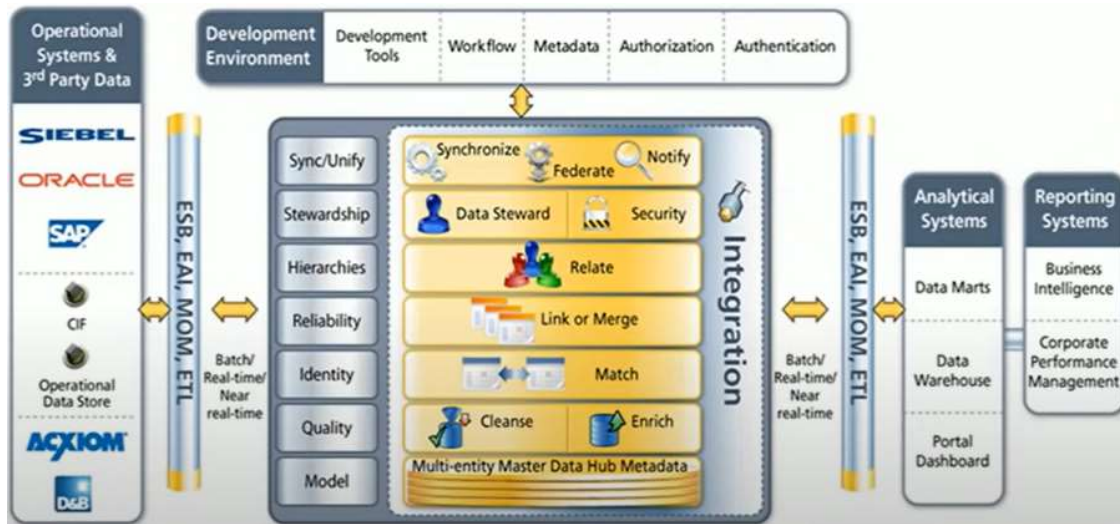


Figura 17. Arquitectura de Referencia de Informática MDM (<https://docs.informatica.com/master-data-management/multidomain-mdm/10-3/overview-guide/informatica-mdm-hub-architecture.html>).

El vendedor Pimcore, aunque se encuentra en el cuadrante “Elección del cliente”, se resalta en color blanco porque su fortaleza está en un MDM especializado para clientes y productos. El vendedor PiLog, aunque también se encuentra en este cuadrante se resalta en color amarillo porque, aunque su plataforma es muy competitiva tecnológicamente, la orientación de la solución no es solo hacia un MDM, sino que también se vende desde otros servicios digitales y por esto es elegido por el cliente.

El vendedor Syndigo clasificado en el cuadrante “Fuerte desempeño” y resaltado en color blanco, tiene la mayor calificación y aun así no fue parte de los vendedores del cuadrante “Elección del cliente”. En el año 2022 si estuvo en ese cuadrante; el cambio posiblemente sucede porque la empresa reformó su base tecnológica y así disminuyó algunas capacidades de atención al cliente; también pudo ser el crecimiento de soluciones como PiLog MDM que logran desplazarlo por su fortaleza en características de gestión de datos entre su front-end y back-end.

El vendedor Reltio clasificado en el cuadrante “Establecido” y resaltado en color amarillo entra como un vendedor con una arquitectura MDM bien definida y atractiva para los clientes. Sin embargo, el número de encuestas diligenciadas por los clientes fue muy poca, entonces lo deja en esta posición al ser una muestra muy pequeña para entrar en un cuadrante como el “Elegido por el cliente”. La figura 20 muestra la arquitectura de referencia MDM de Reltio.

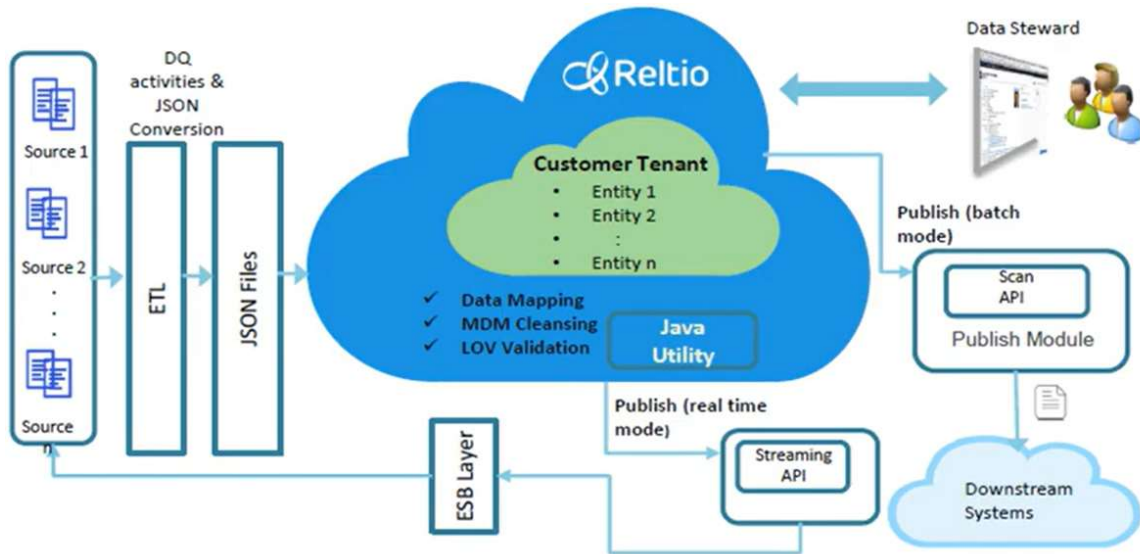


Figura 18. Arquitectura de Referencia MDM – Reltio (<https://www.reltio.com/master-data-management/>).

Los vendedores TIBCO Software e IBM resaltados en color rojo son clasificados en el cuadrante “Aspirantes”. Aunque son vendedores con gran trayectoria en la gestión de datos maestros, posiblemente no fueron bien calificados por sus clientes porque posiblemente la relación front-end / back-end aún no ha evolucionado a ser más flexibles, basados en web u otras características que las nuevas plataformas proveen para agilizar algunas operaciones. La figura 21 muestra la arquitectura de referencia de IBM MDM Cloud.

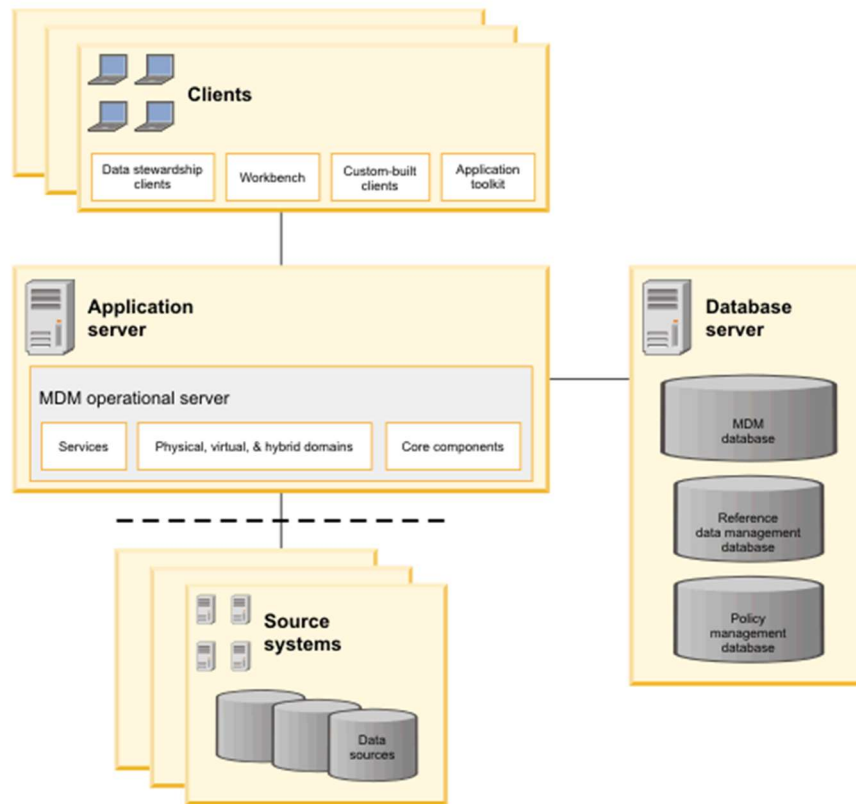


Figura 19. Arquitectura de Referencia MDM – IBM Cloud (<https://www.ibm.com/docs/en/imdm/12.0?topic=overview-architecture-concepts>).

Otros vendedores como Profisee no clasificaron en alguno de los cuadrantes posiblemente porque, aunque tengan arquitecturas robustas apoyadas por proveedores de nube como Azure, aún no tiene un mercado fuerte.

Un factor importante que los clientes también valoran son las categorías y los tipos de licenciamiento. En la tabla 21 se presentan, a modo de ejemplo, las categorías y tipos de licenciamiento ofrecidos por Informatica ya tiene o socios o casos de uso en empresas de Colombia.

Tabla 23. Ejemplo de Categorías y Tipos de Licenciamientos.

Vendedor	Categorías	Tipos
<p>Informatica</p>	<p>Licencias de edición pueden basarse en funciones o en uso.</p> <p>Licencias de conector Las licencias de conector brindan conectividad a entidades como Amazon Redshift, Microsoft SQL Server y Oracle.</p> <p>Licencias personalizadas Las licencias personalizadas son licencias que no forman parte de una edición.</p>	<p>Ensayo (Trial) Puede utilizar la edición de forma gratuita durante un período de 30 días. Al final del período de prueba, puede suscribirse a la edición.</p> <p>Suscripción Puede utilizar la edición con licencia durante el período del contrato.</p> <p>Suscripción gratuita Puede utilizar la tarea de sincronización de forma gratuita</p>

Recomendaciones para el análisis de soluciones para la Gestión de Datos Maestros

- La solución debe proveer una base de datos con componentes de inteligencia artificial para gestionar reglas de calidad básicas y relacionadas con la sincronización e integración de los datos maestros y de referencia.
- La solución debe proveer una base de datos con componentes de inteligencia artificial para gestionar catálogo de datos, gobierno, linaje de datos, seguridad, y protección de datos.
- La solución de proveer interfaces para configurar reglas de calidad, seguridad y protección, así como otras interfaces que faciliten la gestión de todos los componentes definidos en la arquitectura de referencia.
- La solución debe tener un socio ubicado en Colombia para tener un respaldo directo para la implementación de la arquitectura de Referencia.
- El proveedor de la solución debe estar consolidado a nivel mundial y en lo posible tener experiencias de implementación en Colombia.
- El proveedor debe tener su propia nube o estar aliado a nubes como las de AWS, IBM, o Azure.

- Antes de iniciar la fase de implementación, se debe definir la forma más apropiada de extender el modelo de arquitectura de referencia hacia la inclusión de otros dominios/entidades, teniendo en cuenta las siguientes características:
 - Evaluación de la oportunidad de nuevos datos proporcionados por otros dominios para mejorar la toma de decisiones a nivel del Gobierno Nacional.
 - Capacidad de inclusión de nuevos dominios y entidades en los modelos de datos actuales, atributos y otras características asociadas a las estructuras de datos que deban ser modificadas.
 - A nivel de atributos de las entidades maestras o de referencia, los modelos deben ser refinados a medida que las entidades proveedoras cambien o amplíen sus estructuras que almacenan los registros de oro o los datos de referencia.
 - Evaluación de la capacidad del Concentrador GDM para recibir y proveer datos desde y para los nuevos dominios/entidades.
 - Aumentar progresivamente de capacidades en diferentes entidades del Estado para el manejo de la calidad y la seguridad de los datos desde el Concentrador GDM. Esto debe hacerse bajo la implementación de gobernanza definida en el marco de gobierno del Concentrador GDM.
- A nivel de aceptación de la aplicación se deben realizar pruebas de la solución asociadas a los siguientes aspectos:
 - Configuración de roles y responsabilidades.
 - Configuración de permisos y privilegios a los diferentes roles, tipos de usuarios y grupos de usuarios.
 - Monitores de accesos y modificaciones sobre los datos.
 - Configuración de servicios por demanda.
 - Configuración de gestión de la PQRSD para el concentrador GDM, a diferentes temas, ejemplo:
 - Mal uso de datos
 - La calidad de los datos
 - Modificación de datos
 - Problemas de acceso
 - Entre otros PQRSD
- A nivel de seguridad y privacidad, al momento de hacer pruebas sobre la arquitectura, se deben incluir:
 - Pruebas de penetración y vulnerabilidad para identificar posibles riesgos de seguridad en la periodicidad de ejecución. En otras palabras, este tipo de pruebas permitirán identificar los diferentes fallos o errores de seguridad que pueden afectar los datos, aplicaciones, redes, etc., que puedan afectar la gestión de los datos maestros del Concentrador GDM. Estas pruebas se deben hacer en los siguientes niveles:
 - Configuración de roles y responsabilidades.
 - Configuración de permisos y privilegios a los diferentes roles, tipos de usuarios y grupos de usuarios.

- Monitoreo de accesos y modificaciones sobre los datos para final se debe monitorear y hacer seguimiento de eventos de seguridad para detectar y prevenir posibles incidentes de seguridad.
- A nivel de Recomendación –Restricciones de la Arquitectura: Acuerdos de Servicio. El artículo 10 de la Ley 2106 de 2019 establece lo siguiente:
 - "Artículo 10. No se requiere acuerdo de servicio. No será necesario el acuerdo de servicio, contrato o autorización del titular para el tratamiento de datos personales cuando se trate de:
 - Información requerida por una entidad pública o administrativa en ejercicio de sus funciones legales o por orden judicial.
 - Datos de naturaleza pública.
 - Casos de urgencia médica o sanitaria.
 - Tratamiento de información autorizado por la ley para fines históricos, estadísticos o científicos, previa disociación de los datos personales."
 - El artículo establece que en ciertos casos específicos no se requiere el acuerdo de servicio, contrato o autorización del titular para el tratamiento de datos personales. Dichos casos incluyen la información requerida por una entidad pública o administrativa en ejercicio de sus funciones legales o por orden judicial, datos de naturaleza pública, casos de urgencia médica o sanitaria, y tratamiento de información autorizado por la ley para fines históricos, estadísticos o científicos previa disociación de los datos personales.
- A nivel de pruebas de estrés, estas se deben definir conjuntamente entre el Administrador de Datos y el proveedor de la solución tecnológica. Se debe tener en cuenta la medición de la solidez y las capacidades de manejo de errores las aplicaciones y manejo del motor de la base de datos a nivel de condiciones de carga, y de extracción desde los procesos de ingesta de datos desde las entidades proveedoras. De esta forma las pruebas deben garantizar que el sistema no se bloquee en situaciones críticas.

Glosario

Auditoría de Datos. “Es un medio para detectar el acceso no autorizado a los datos o las aplicaciones codificadas de forma insegura.” (IBM, 2023).

Back-end – Front-end. “En la ingeniería de software, los términos front-end y back-end (o a veces denominados back-end o back-end) se refieren a la separación de intereses entre la capa de presentación (front-end) y la capa de acceso a datos (back-end) de una pieza de software. o la infraestructura física o el hardware. En el modelo cliente-servidor, el cliente generalmente se considera el frontend y el servidor generalmente se considera el backend, incluso cuando parte del trabajo de presentación se realiza en el propio servidor.” (Wikipedia, 2023).

Bases de Datos orientadas a Grafos (BDOG). “Este tipo de bases de datos representa la información como nodos de un grafo y sus relaciones con las aristas de este, de manera que se pueda usar teoría de grafos para recorrer la base de datos ya que esta puede describir atributos de los nodos (entidades) y las aristas (relaciones).” (Wikipedia, 2023).

Canalizador (pipeline). Una canalización se puede considerar una serie de tareas detalladas. Es una cadena de procesos conectados de forma tal que la salida de cada elemento de la cadena es la entrada del próximo. (Wikipedia, 2023).

Clave Criptográfica. “Una clave en criptografía es una pieza de información, generalmente una cadena de números o letras que se almacenan en un archivo que, cuando se procesa a través de un algoritmo criptográfico, puede codificar o decodificar datos criptográficos. Según el método utilizado, la clave puede tener diferentes tamaños y variedades, pero en todos los casos, la solidez del cifrado depende de la seguridad de la clave que se mantiene. La solidez de la seguridad de una clave depende de su algoritmo, el tamaño de la clave, la generación de la clave y el proceso de intercambio de claves.” (Wikipedia, 2023).

Cliente huella cero. “Un cliente de huella cero significa que no hay instalación o descarga del lado del cliente, lo que permite la visualización de documentos e imágenes dentro del navegador web nativo, mientras aprovecha las verdaderas capacidades del navegador, los complementos integrados y su interacción con el propio dispositivo.” (RiskWatch, 2023).

Computación bajo demanda (On-Demand Computing). “Es un modelo de entrega en el que los recursos informáticos se ponen a disposición del usuario según sea necesario. Los recursos pueden mantenerse dentro de la empresa del usuario o ser puestos a disposición por un proveedor de servicios en la nube. El término computación en la nube a menudo se usa como sinónimo de computación bajo demanda cuando los servicios son proporcionados por un tercero, como una organización de alojamiento en la nube. El modelo de computación comercial bajo demanda se desarrolló debido a que la demanda de recursos informáticos de una empresa puede ser impredecible en ocasiones, mantener suficientes recursos para cumplir con los requisitos máximos puede resultar costoso. Y reducir los costos manteniendo solo los recursos mínimos significa que es

probable que no haya suficientes recursos para cumplir con las cargas máximas. El modelo bajo demanda brinda a una empresa la capacidad de escalar los recursos informáticos hacia arriba o hacia abajo cuando sea necesario, con solo hacer clic en un botón. El modelo se caracteriza por tres atributos: escalabilidad, pago por uso y autoservicio. Ya sea que el recurso sea un programa de aplicación que ayude a los miembros del equipo a colaborar o proporcione almacenamiento adicional, los recursos informáticos son elásticos, medidos y fáciles de obtener.” (Gillis, 2023).

CRUD. Sigla que representa las cuatro operaciones que se realizan sobre los datos: Create, Retrieve, Update, y Delete,

Entrada (Check-in) y Salida (Check-out) de una licencia. “Una vez que se hace check-out de una licencia, no está disponible para otros usuarios potenciales hasta que la aplicación la devuelva al salir, o hasta que se agote el tiempo de espera de la licencia. Por lo general, hay un número máximo de licencias simultáneas que se pueden retirar en cualquier momento y, por lo tanto, la solicitud de retirar una licencia puede ser denegada si se produce esta situación (denegación de servicio).” (Zendtitle, 2023).

ETL. Sigla que representa los procesos de Extract, Transform, Load. Estos procesos se hace sobre los datos que pasa de ser transaccionales a ser datos para la toma de decisiones o para la gobernanza de una organización.

Fuente externa. Fuente externa es una organización que produce datos a nivel local, nacional o mundial que son de interés del gobierno Colombia para ser incluidos en el Concentrador GDM como datos de referencia.

Hooking. Hooking es un término de programación informática que se refiere a una colección de técnicas empleadas para cambiar el comportamiento de las aplicaciones o los sistemas operativos. El enlace implica la intercepción de llamadas a funciones, eventos del sistema o mensajes, y los fragmentos de código que realizan estas intercepciones se denominan enlaces. (VMRay, 2023).

Ingesta de Datos. “El proceso de ingesta generalmente requiere una secuencia de operaciones, desde recuperar los datos hasta analizarlos, validarlos, transformarlos y enriquecerlos, hasta cargarlos y archivarlos.” (CloverDX, 2023). Actualmente se han declarado tres tipos de ingesta de datos: en tiempo .real, en lote, y combianando de los dos anteriores (striim, 2023).

JSON. “JavaScript Object Notation, es un formato de archivo estándar abierto y un formato de intercambio de datos que utiliza texto legible por humanos para almacenar y transmitir objetos de datos que consisten en pares de atributos-valores y matrices (u otros valores serializables). valores).” (Wikipedia, 2023)

Licencias de rango corto. Se refiere a licencias que se pagan por una sola vez o que están limitadas a rangos de tiempos de uso de una licencia.

Orquestar Plataforma. Se define como una tarea del ambiente de operación de los datos. Esta “involucra la creación de máquinas virtuales o contenedores, varios procesos de tiempo de ejecución basados en el contexto

de la tarea de datos en cuestión, procesos de transferencia y almacenamiento de datos y monitoreo. El escaneo automatizado de SQL u otras bases de datos ayuda a identificar identificaciones incorrectas u otros parámetros que pueden indicar datos no deseados o inexactos antes de que lleguen a su plataforma de análisis. La orquestación de infraestructura automatizada ayuda a aprovisionar recursos adicionales para manejar una gran afluencia de datos.” (LUNAVI, 2023).

Registro de Oro. “Es una versión única y bien definida de todas las entidades de datos en un ecosistema organizacional. A veces se denomina la "versión única de la verdad", entendiendo por "verdad" la referencia a la que los usuarios de datos pueden recurrir cuando quieren asegurarse de que tienen la versión correcta de un fragmento de información. información.” (Whats.com, 2023).

REST. “es un estilo de arquitectura software para sistemas hipermedia distribuidos como la World Wide Web.” (Wikipedia, 2023). API REST, específicamente

Servicios medidos o pago por uso. “Los servicios medidos, también llamados pago por uso y precios basados en el consumo, son cualquier tipo de estructura de pago en la que un cliente tiene acceso a recursos potencialmente ilimitados, pero solo paga por lo que realmente usa. Los servicios medidos son cada vez más comunes en los entornos de TI empresariales.” (Hanna, 2023),

Servicio Web. “es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet.” (Wikipedia, 2023).

SOAP. “es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos XML.”

SQL. Sigla que representa Structure Query Language. Es un lenguaje de dominio específico, diseñado para administrar, y recuperar información de sistemas de gestión de bases de datos relacionales. (Wikipedia, 2023).

Tokenización. “La tokenización, cuando se aplica a la seguridad de los datos, se refiere al proceso de sustitución de un elemento de datos sensible por un equivalente no sensible denominado token, que no tiene un significado o valor extrínseco o explotable. El token es una referencia (un identificador) que regresa a los datos sensibles a través de un sistema de tokenización.” (Wikipedia, 2023).

TXT. “Los archivos TXT son documentos de texto ASCII sin formato, también llamados "texto sin formato". Fuera de los países de habla inglesa, TXT también puede significar texto extendido.” (Wiki_Mobileread, 2023).

XML. “Lenguaje de marcado extensible (Extensible Markup Language) es un lenguaje de marcado y un formato de archivo para almacenar, transmitir y reconstruir datos arbitrarios. Define un conjunto de reglas para codificar documentos en un formato que sea tanto legible por humanos como por máquinas.” (Wikipedia, XML, 2023)

Referencias

- 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).
- Agency for Data Supply and Infrastructure. (2022). *The Danish Basic Data Program, the Data-Distributor (Datafordeleren) and the Covid-19 Pandemic*. Genova.
- Allen, M., & Cervo, D. (2015). *Multi-Domain Master Data Management*. Morgan Kaufmann.
- Alsarkhi, A., & Talburt, J. R. (2018). A Method for Implementing Probabilistic Entity Resolution. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 9(11).
- BPMN. (27 de 3 de 2023). *Object Management Group usiness Process Model and Notation*. Obtenido de <https://www.bpmn.org/>
- Buyya, R., Vecchiola, C., & Selvi, S. T. (2013). *Cloud computing Bible & Mastering Cloud computing*. Morgan Kaufmann.
- 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>
- 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>
- 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.
- Cloutier, R., Muller, G., Verma, D., & Nilchiani, R. (2008). *The Concept of Reference Architectures*. Wiley InterScience.
- CloverDX. (31 de 3 de 2023). *Data validation in data ingestion processes*. Obtenido de <https://www.cloverdx.com/blog/data-validation-in-data-ingestion-processes>
- 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>
- ComputerWeekly.es. (3 de 3 de 2023). *Autenticación Multifactor*. Obtenido de <https://www.computerweekly.com/es/definicion/Autenticacion-multifactor-o-MFA>
- DAMA Internacional. (10 de 10 de 2022). *DAMA Internacional*. Obtenido de <https://www.dama.org/cpages/home>

- DAMA International. (2017). *DMBOK: Data Management Body of Knowledge*. Basking Ridge, New Jersey: Technics Publications.
- 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 - Guía Diseño RA AEI. (2021). *GUÍA PARA EL DISEÑO DE REGISTROS ADMINISTRATIVOS DE ACTIVIDADES ECONÓMICAS INFORMALES*.
- DANE - Guía para carga a las fuentes. (2020). *GUÍA PARA LA REDUCCIÓN DE LA CARGA A LAS FUENTES*.
- DANE - Informe Estadística Sociodemográfica. (2022). *Informes de Estadística Sociodemográfica Número 14 - Principales ciudades y sus conurbaciones en Colombia*. Bogotá: DANE.
- DANE - METODOLOGIA REG ADMIN. (2021). *METODOLOGÍA GENERAL PARA LA CONSTRUCCIÓN Y ACTUALIZACIÓN DEL REGISTRO ESTADÍSTICO BASE DE POBLACIÓN – REBP 2020*.
- 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 - 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)*.
- 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>
- DANE. (2018). *Guía de Estándares para Registros Administrativos*.
- 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>
- DANE Conceptos estandarizados. (10 de 10 de 2022). *SISTEMA DE CONSULTA DE CONCEPTOS ESTANDARIZADOS*. Obtenido de <https://conceptos.dane.gov.co/conceptos/transversal>
- DANE GIT DEST. (2022). *Sistema de Registros Estadísticos*.
- DANE-RegAdmitivos. (12 de 10 de 2022). <https://www.dane.gov.co/index.php/sistema-estadistico-nacional-sen/registros-administrativos/programa-de-fortalecimiento>.
- Danish Government; Local Government Denmark. (2012). *Good Basic Data for Everyone – A Driver for Growth and Efficiency*. Denmark.
- Denmark, E. C. (2021). Digital Public Administration factsheet 2021 - Denmark.
- 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
- Departamento Administrativo de la Función Pública. (2012). *Ley 1581 de 2012 - Protección de Datos Personales*.
- DNP - Comité Política Catastro Multipropósito. (2022). *Comité de Implementación de la Política de Catastro Multipropósito (ppt)*. Bogotá.
- DNP - CONPES 3958. (2019). *CONPES 3958 - ESTRATEGIA PARA LA IMPLEMENTACIÓN DE LA POLÍTICA PÚBLICA DE CATASTRO MULTIPROPÓSITO*. Bogotá.
- DNP. (2012). Bogotá.

- DNP. (2017). *Análisis de duplicidad en la información recolectada por medio de Registros Únicos*. 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>
- 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; 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-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.
- eInforma. (27 de 12 de 2022). *Directorio de Empresas de Colombia*. Obtenido de <https://directorio-empresas.einforma.co/informacion-empresa/importadora-bombonera-he-sas>
- Entur. (30 de 8 de 2022). *Stops- and Timetable data*. Obtenido de <https://developer.entur.org/stops-and-timetable-data>
- European Commission - Estonia. (2022). *Digital Public Administration factsheet 2022 - Estonia*.
- European Commission - Poland. (2022). *Digital Public Administration factsheet 2022 Poland*.
- Eurostat. (2010). *Manual de recomendaciones de Registros de Empresas*.
- 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".
- Gamero, A., Garcia, 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).
- Gartner Peer Insights. (10 de 02 de 2023). *Gartner Peer Insights Voice of the Customer Methodology*. Obtenido de <https://blogs.gartner.com/reviews-pages/gartner-peer-insights-voice-customer-methodology-2-0/>
- Gartner Peer Insights R&R. (12 de 2 de 2023). *Master Data Management Solutions Review and Ratings*. Obtenido de <https://www.gartner.com/reviews/market/master-data-management-solutions>
- Gillis, A. S. (17 de 3 de 2023). *on-demand computing*. Obtenido de <https://www.techtarget.com/searchitoperations/definition/on-demand-computing>
- 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>
- 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
- Hanna, K. T. (17 de 3 de 2023). *metered services (pay-per-use)*. Obtenido de <https://www.techtarget.com/searchcio/definition/metered-services>
- Helleland, L. H. (2021). *Norwegian data centres - sustainable, digital powerhouses*. Norwegian Ministry of Local Government and Modernisation.

- Horst, N. L., Bjerre, S., Lind, M., & Hvingel, L. (2014). The Basic Data Programme – A Danish Infrastructure Model for Public Data. *PERSPEKTIV NO. 24*.
- ICDE. (10 de 10 de 2022). *ICDE - Datos Fundamentales*. Obtenido de <https://www.icde.gov.co/datos-y-recursos/datos-fundamentales>
- 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>
- IGAC - Avances Catastro Multipropósito. (2021). *Avances en la implementación de la política de Catastro Multipropósito*.
- IGAC - Portal. (2022). *IGAC - Que estamos haciendo?* Obtenido de <https://www.igac.gov.co/es/contenido/areas-estrategicas/catastro/que-estamos-haciendo>
- IGAC - Reporte de Información Catastral. (2022). *REPORTE DE INFORMACIÓN CATASTRAL*. Obtenido de <https://igac.gov.co/catastro-multiproposito/reporte-de-informacion-catastral>
- 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. (2022). *IGAC - LADM_COL*. Obtenido de <https://igac.gov.co/es/catastro-multiproposito/ladm-col>
- 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>
- ISO. (2012). *ISO standar 19152*. Obtenido de Geographic information — Land Administration Domain Model (LADM): <https://www.iso.org/standard/51206.html>
- ISO. (12 de 2 de 2023). *ISO 25000 Calidad de Software y Datos*. Obtenido de <https://iso25000.com/index.php/normas-iso-25000/iso-25010/23-usabilidad>
- 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>
- Jonker, R., Kooistra, F., Cepariu, D., Etten, J. v., & Swartjes, S. (2011). *Effective master data management*. KPMG.
- Karan, E. P., Irizarry, J., & Haymaker, J. (2015). *BIM and GIS integration and interoperability based on semantic web technology*.
- Kaspersky. (26 de 3 de 2023). *¿Qué es el cifrado de datos? Definición y explicación*. Obtenido de <https://latam.kaspersky.com/resource-center/definitions/encryption>
- Kendrick, T. (2013). *The Project Management Tool Kit, 3rd Edition*. AMACOM.
- 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/>
- 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).

- 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).
- Kroon, J., & Paas, A. (2014). *Using registers as the basis to statistical production – linking issues and potential bottlenecks*. IFC Bulletin No 37.
- Ligus, S. (2012). *Effective Monitoring and Alerting*. O'Reilly.
- Ministerio de Industria y Comercio. (2013). *Decreto Régimen General de Protección de Datos*.
- Ministry of Local Government and Modernisation. (2019). *One Digital Public Sector - Digital strategy for the public sector 2019–2025*. Oslo.
- MinSalud. (2022). *ADRES*. Obtenido de <https://www.adres.gov.co/entidades-territoriales/bdua>
- MinSalud. (2022). *Planilla Integrada de Liquidación de Aportes*. Obtenido de <https://www.minsalud.gov.co/proteccionsocial/Paginas/pila.aspx>
- 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
- MinTIC - Marco de interoperabilidad. (2019). *Marco de interoperabilidad para Gobierno Digital*.
- 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>
- 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
- MinTIC - SCD. (2020). *Anexo 2 Guía para la Vinculación y Uso de los Servicios Ciudadanos Digitales*. Bogotá: MinTIC.
- 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
- 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
- Minvivienda. (2021). *MARCO DE REFERENCIA DE INTEGRACIÓN E INTERCAMBIO DE INFORMACIÓN*. Ministerio de Vivienda.
- Nacer, A. (2021). *Gobernanza digital e interoperabilidad gubernamental: una guía para su implementación*. Documentos de Proyectos (LC/TS.2021/80) . Santiago: Comisión Económica para América Latina y el Caribe (CEPAL).
- 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
- 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>
- 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>
- NordREG. (2021). *Implementation of data hubs in the Nordic countries*. Oslo: Nordic Electric Regulators.
- Norway, E. C. (2020). *Digital Government Factsheet 2019 Norway*.

- Norway, European Commission. (2022). *Digital Public Administration 2021*.
- 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.
- 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>
- Pimcore. (20 de 02 de 2023). *The Leading Master Data Management Solutions*. Obtenido de <https://pimcore.com/en/resources/gartner-reports/peer-insights-voc-mdm-2022>
- Plotkin, D. (2014). *Data Stewardship - An Actionable Guide to Effective Data Management and Data Governance*. Elsevier Inc.
- 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
- 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)*.
- 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
- Registraduria . (2022). *ACCESO A LA BASE DE DATOS ANI*. Obtenido de <https://www.registraduria.gov.co/-Consultas-ANI-.html>
- 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.
- 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.
- RiskWatch. (17 de 3 de 2023). *The Benefits of a Zero Footprint Viewer*. Obtenido de <https://riskwatch.com/2012/07/25/the-benefits-of-a-zero-footprint-viewer/>
- 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.
- 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>
- striim. (31 de 3 de 2023). *What is Data Ingestion and Why This Technology Matters*. Obtenido de <https://www.striim.com/blog/what-is-data-ingestion-and-why-this-technology-matters/>
- TIBCO. (25 de 2 de 2023). <https://www.tibco.com/resources/analyst-report/the-forrester-wave-master-data-management>. Obtenido de <https://www.tibco.com/resources/analyst-report/the-forrester-wave-master-data-management>
- Vassil, K. (2015). *Estonian e-Government Ecosystem - Foundation, applications, outcomes*. Institute of Government and Politics - University of Tartu.
- VMRay. (17 de 03 de 2023). *Hooking*. Obtenido de <https://www.vmrays.com/glossary/hooking/#:~:text=Hooking%20is%20a%20computer%20programming,these%20interceptions%20are%20called%20hooks>.

- Wallgren, Wallgren, A., & Britt, y. (2021). *Hacia un sistema estadístico integrado y basado en registros*. New York: BID.
- 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>
- Wikipedia. (17 de 3 de 2023). *Bases de datos orientadas a grafos*. Obtenido de https://es.wikipedia.org/wiki/Base_de_datos_orientada_a_grafos
- Wikipedia. (17 de 3 de 2023). *Frontend - Backend*. Obtenido de https://en.wikipedia.org/wiki/Frontend_and_backend
- Wikipedia. (25 de 3 de 2023). *Servicio Web*. Obtenido de https://es.wikipedia.org/wiki/Servicio_web
- Wikipedia. (27 de 3 de 2023). *Transferencia de Estado Representacional*. Obtenido de https://es.wikipedia.org/wiki/Transferencia_de_Estado_Representacional
- Woody, A. (2013). *Enterprise Security: A Data-Centric Approach to Securing the Enterprise*. Packt Publishing.
- Zendtitle. (17 de 3 de 2023). *Check-out*. Obtenido de <https://www.nalpeiron.com/glossary/check-out/>

Anexo 1. Dimensiones y Reglas de Calidad de los Datos

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**.

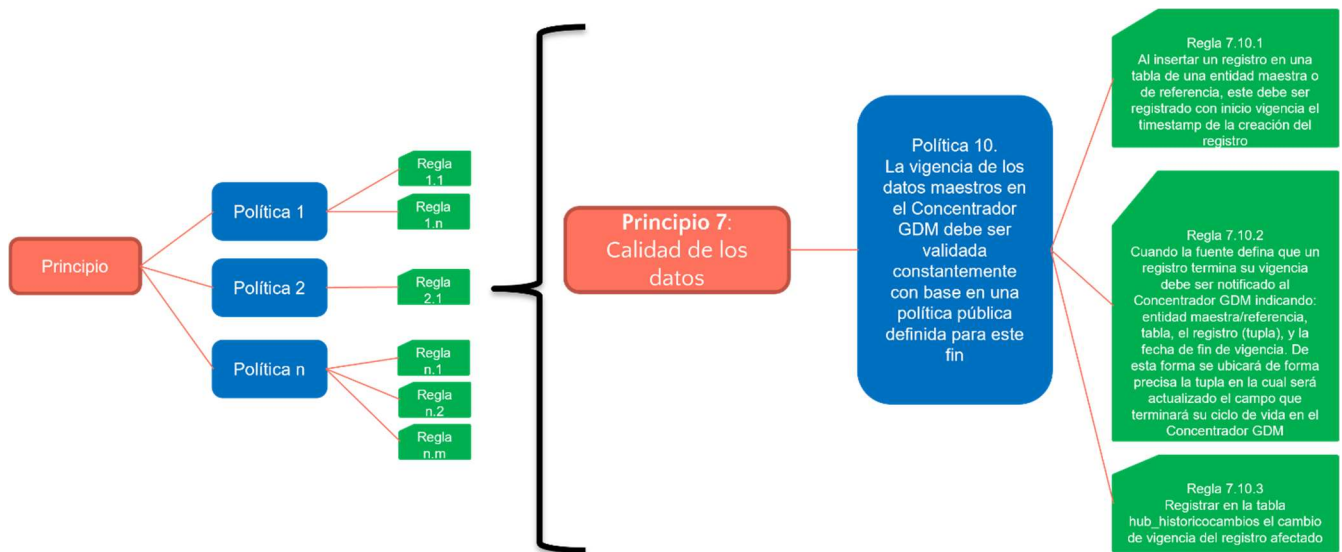


Figura 20. Jerarquía para el aseguramiento de la calidad de los datos maestros y de referencia.

8.1.1. 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”.

8.1.2. 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.

8.1.3. 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.

8.1.4. 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.

8.1.5. 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.

8.1.6. 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.

8.1.7. 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.

8.1.8. 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)

8.1.9. 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.

8.1.10. 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.

8.2. 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

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

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.

8.3. 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:

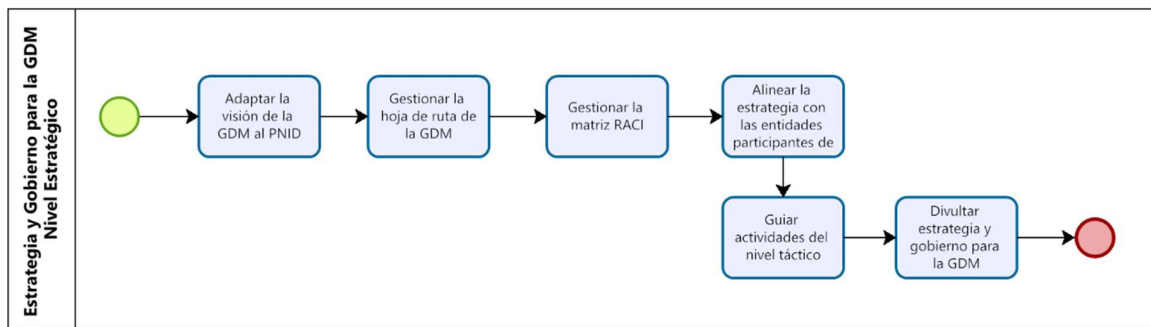
- **Más reciente:** se pueden ordenar los registros con sello de fecha en orden del más al menos reciente.
- **Más frecuente:** coincide con registros repetidos que contienen la misma información
- **Más completo:** considera que el campo está completo como una indicación de la precisión del registro.

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.

Anexo 2. Proceso Estrategia y Gobierno para la GDM Nivel Estratégico



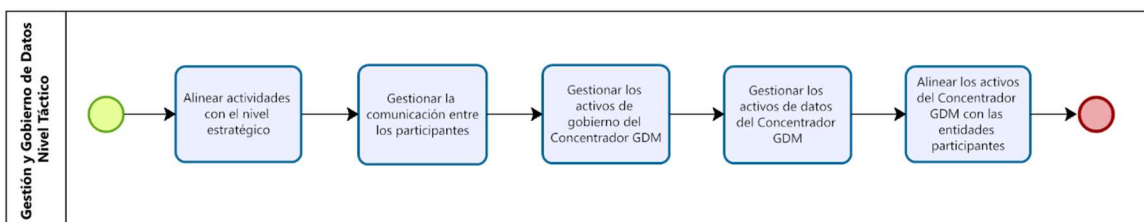
Powered by
bizagi
Modeler

Figura 21. Proceso Estrategia y Gobierno para la GDM Nivel Estratégico.

Actividades del proceso:

- **Inicio.** El proceso inicia cuando el equipo del nivel estratégico requiere tomar decisiones a nivel de estrategia o gobierno de los datos maestros o de referencia.
- **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.
- **Fin.** El proceso finaliza cuando el equipo estratégico termina un ciclo de deliberación y toma de decisiones a nivel de estrategia o el gobierno.



Powered by
brazgi
Modeler

Figura 22. Proceso Gestión y Gobierno de Datos Nivel Táctico.

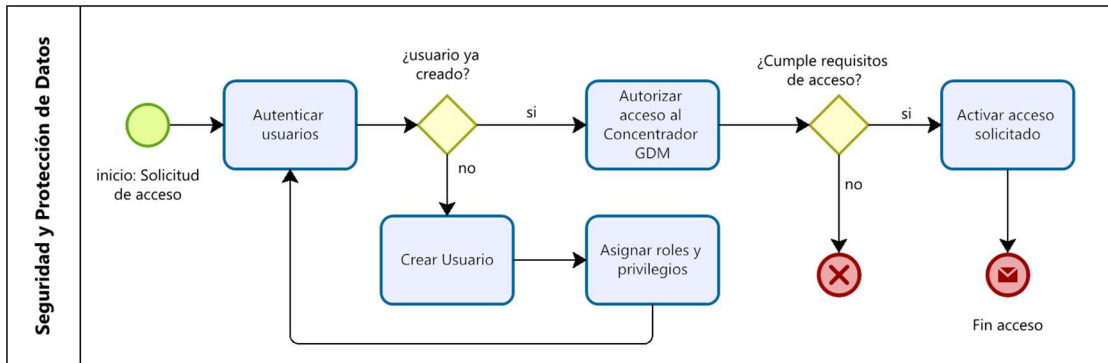
Actividades del proceso:

- **Inicio.** El proceso inicia cuando el equipo del nivel táctico requiere tomar decisiones a nivel de gestión o gobierno de los datos maestros o de referencia.

- **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 tiempo⁸. 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.
- **Fin.** El proceso finaliza cuando el equipo táctico termina un ciclo de deliberación y toma de decisiones a nivel de gestión o el gobierno.

⁸ Esta actividad es una de las responsabilidades del administrador de datos, de acuerdo con el decreto 1389 de 2022 (Función Pública, 2022).

Anexo 3. Proceso de Seguridad y Protección de los datos



Powered by
bizagi
Modeler

Figura 23. Proceso Seguridad y Protección de Datos.

Actividades del Proceso

- **Inicio: Solicitud de acceso.** El proceso inicia cuando alguno de los roles definidos a nivel del gobierno del Concentrador GDM desea ingresar al sistema con un usuario y contraseña.
- **Autenticar usuarios.** Esta tarea verifica la identidad del usuario y contraseña en el Concentrador GDM. Genera un estado de existencia o no del usuario en el sistema.
- **Decisión: ¿usuario ya creado?** Recibe un estado: Si o No está autenticado.
 - **Si** está creado. Activa la tarea “Autorizar acceso al Concentrador GDM”
 - **No** está creado. Activa la tarea “Crear Usuario”
- **Autorizar acceso al Concentrador GDM.**
- **Crear Usuario.** Se procede a crear el usuario y contraseña en el Concentrador GDM.
- **Asignar roles y privilegios.** Una vez creado el usuario y su contraseña se asignan roles y privilegios. Esto activa de nuevo la autenticación para que el nuevo usuario sea autorizado de acuerdo con su solicitud de acceso. Genera un estado: Si o no cumple requisitos de acceso
- **Decisión: ¿Cumple requisitos de acceso?** Recibe un estado: Si o No cumple requisitos de acceso.
 - **Si** cumple. Activa la tarea “Activar acceso solicitado”
 - **No** cumple. Activa el evento de Fin de Cancelación, el cual inhibe la continuidad de las acciones solicitadas sobre el Concentrador GDM
- **Activar acceso solicitado.** El usuario es autorizado en su lista de privilegios: aplicaciones, datos, estructuras, etc.

Fin acceso. Activa el evento de fin del proceso.