



GUIA GENERAL

MGGTI.G.GI - DOMINIO DE INFORMACIÓN

Ministerio de Tecnologías de la Información y las Comunicaciones 2023

MGGTI

Ministerio de Tecnologías de la Información y las Comunicaciones

Viceministerio de Transformación Digital

Subdirección de Estándares y Arquitectura de Tecnologías de la Información

Equipo de trabajo

Mauricio Lizcano Arango - Ministro de Tecnologías de la Información y las Comunicaciones

Sindey Carolina Bernal Villamarín - Viceministra de Transformación Digital

Ana María Sterling Bastidas – Directora de Gobierno Digital

Luis Clímaco Córdoba Gómez - Subdirector de Estándares y Arquitectura de TI

Jairo Alberto Riascos Muñoz – Equipo de la Subdirección de Estándares y Arquitectura de TI

Claudia Milena Rodríguez Alvarez- Equipo de la Subdirección de Estándares y Arquitectura de TI

Luis Martin Barrera Pino - Equipo de la Subdirección de Estándares y Arquitectura de TI

Samuel Antonio Peña Navarro - Equipo de la Subdirección de Estándares y Arquitectura de TI

Empresa Consultora Yobiplex

Versión

Observaciones

Versión 1.0
Mayo 2023

Guía General Dominio de Gestión de Información

Tabla de contenido

Listado de ilustraciones	4
Listado de tablas.....	5
1. Introducción	6
1.1. Usted Está Aquí.....	8
1.2. Propósito de esta guía	9
1.3. A quién va dirigida	10
2. Lineamientos	11
3. Etapas	13
3.1. Gestión del Ciclo de vida de los datos	14
3.2. Gobierno de datos e Información.....	14
3.3. Calidad de datos	14
3.4. Identificación y gestión de Datos maestros	15
3.5. Identificación y Gestión de Metadatos	16
Catálogo de Metadatos	16
3.6. Explotación de los datos	17
3.7. Apertura de datos.....	18
3.8. Gestión de Documentos electrónicos	18
3.9. Identificación y gestión de datos geoespaciales	19
4. Roles	21
5. Caso práctico	24
5.1. Contexto25	
5.2. Situación	26
5.3. Primeros pasos	27
5.4. Pasos adicionales.....	28
6. Artefactos	30
7. Estándares y Mejores prácticas	32

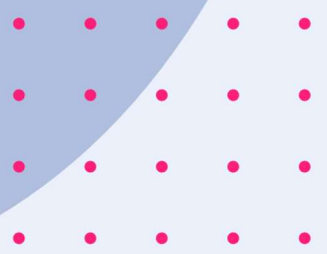
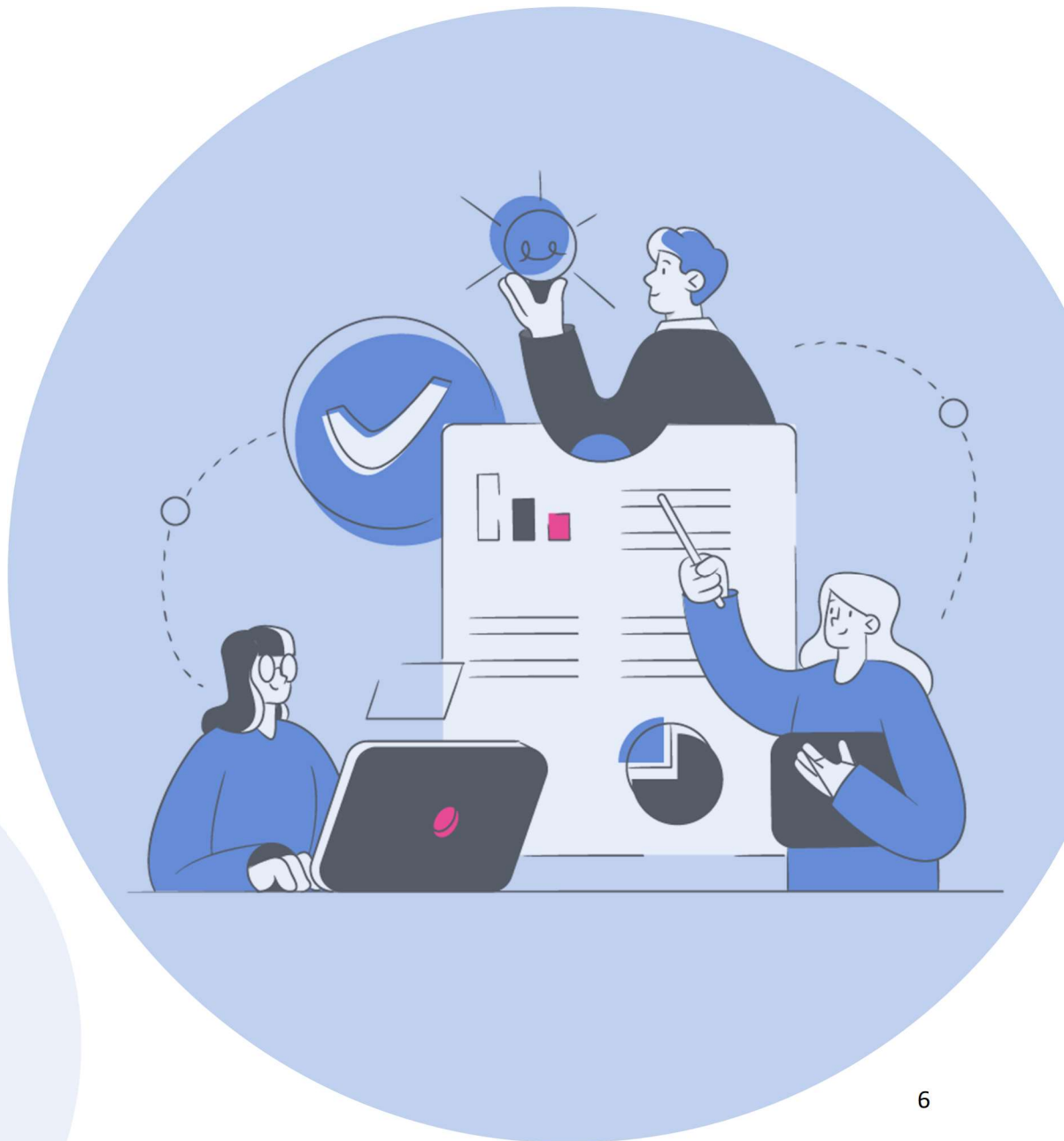
Listado de ilustraciones

Ilustración 1. Pirámide del conocimiento.....	7
Ilustración 2. Pirámide del conocimiento – Ejemplo trivial.....	8
Ilustración 3. Ubicación de la Guía de Gestión de Información en el contexto del Modelo de Gestión y Gobierno de TI	9
Ilustración 4. Audiencia	10
Ilustración 5. Proceso para gestionar Metadatos	16
Ilustración 6. Catálogo de Metadatos	17
Ilustración 7. Esquema de la estructura de sedes dependientes de instituciones educativas	25
Ilustración 8. Esquema de la estructura de edificaciones dependiendo de sedes educativas	26

Listado de tablas

Tabla 1. Lineamientos del Dominio de Gestión de Información	12
Tabla 2. Roles de la Arquitectura de Información	23
Tabla 3. Lista de artefactos.....	31
Tabla 4. Estándares y mejores prácticas	34

1. Introducción



Desde que se tiene noción de la gestión de los datos como disciplina, ha existido la necesidad de definir y diferenciar los conceptos: datos, información, conocimiento y, posteriormente, sabiduría. El razonamiento más aceptado se describe gráficamente a través de una pirámide en la que cada nivel corresponde a una definición y entre más arriba de la pirámide se esté más valor se genera; la siguiente ilustración describe dicha pirámide:

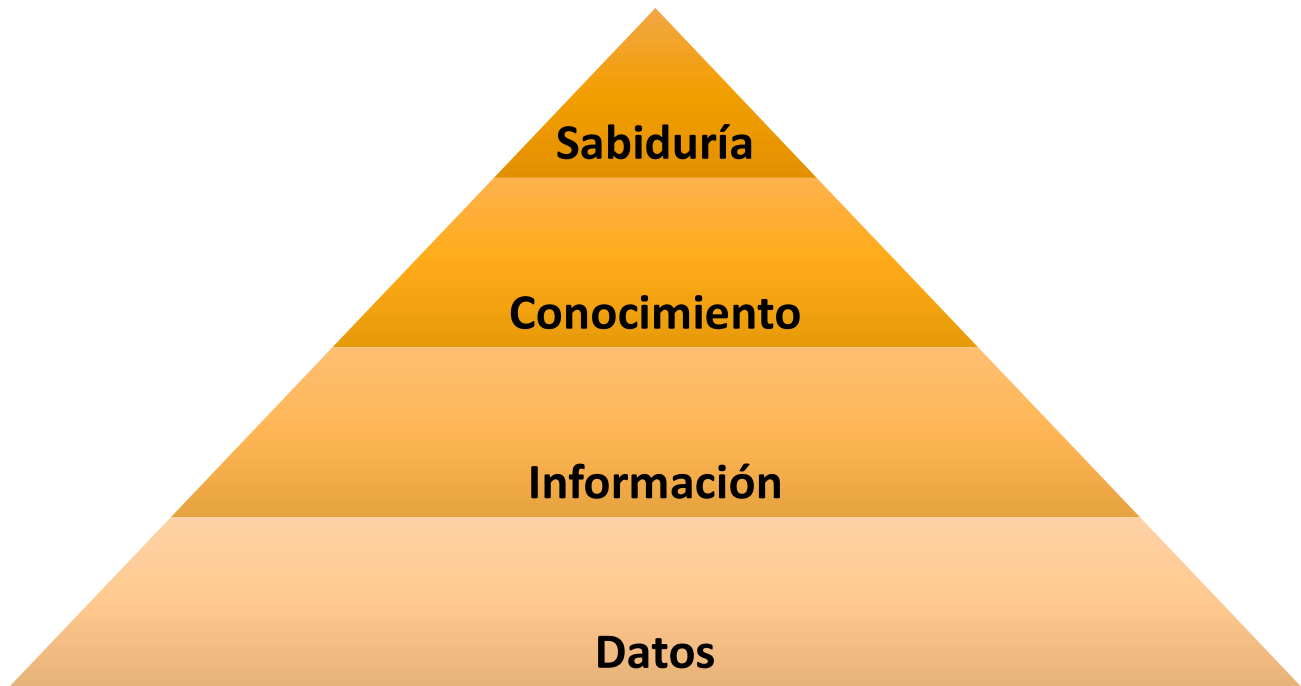


Ilustración 1. Pirámide del conocimiento

La matriz asume que existen los datos y que los datos puestos en un contexto específico generan información; asimismo, cuando se coloca la información en contexto para lograr un fin tengo conocimiento y finalmente la sabiduría se obtiene a través del conocimiento aplicado a acciones concretas. La siguiente pirámide describe un caso trivial, para tratar de facilitar el entendimiento de los conceptos:

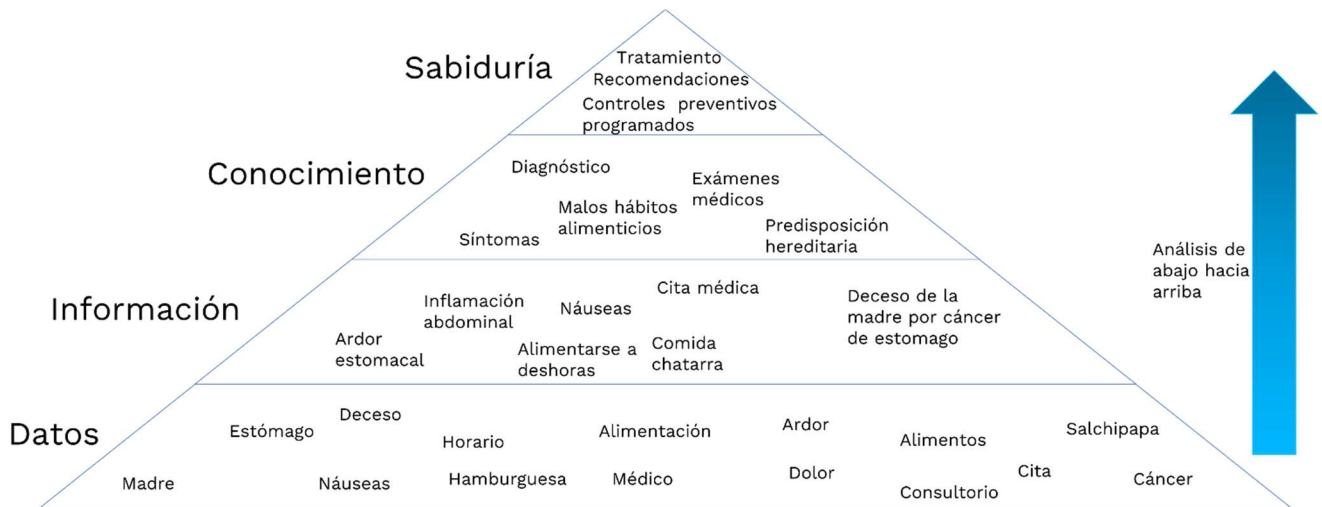


Ilustración 2. Pirámide del conocimiento – Ejemplo trivial.

Elaboración propia.

Se puede observar en la ilustración, que en el primer nivel hay una serie de datos que sin ningún tipo de contexto carecen de significado, cuando se relacionan entre sí esos datos, se genera información; cuando esta información llega al médico, este genera información adicional en base a esa información primaria y con todo ese contexto genera un diagnóstico, identifica una predisposición y malos hábitos. Con el conocimiento del médico y la información gestionada el médico define un tratamiento para curar la enfermedad y dicta una serie de recomendaciones, la sabiduría en este caso está dada por las acciones que el médico define en base a su conocimiento del caso de la gestión de la información.

Los datos y la información dependen el uno del otro y están asociados al pasado (hechos), mientras que el conocimiento y la sabiduría se asocian al presente y futuro (perspectivas). A primera vista el conocimiento y la sabiduría se roban toda la atención; sin embargo, sin datos e información de calidad y adecuadamente gestionados no es posible llegar a los niveles más altos de la pirámide (conocimiento y sabiduría). Las entidades del Estado colombiano deben gestionar adecuadamente su información para poder obtener conocimiento y sabiduría que les permita cumplir a cabalidad con su misión, esta guía resume los puntos más importantes para tener en cuenta para gestionar información.



(Tenga en cuenta) Las guías tienen un carácter orientador y no son de obligatoria aplicación (ARTÍCULO 2.2.9.1.2.2. Decreto 767 de 2022. Política de Gobierno digital). Esta guía busca orientar a las entidades públicas para definir su estrategia de TI.

1.1. Usted Está Aquí

En la siguiente ilustración se muestran los principales componentes del dominio de gestión de información, perteneciente al Modelo de Gestión y Gobierno de TI.

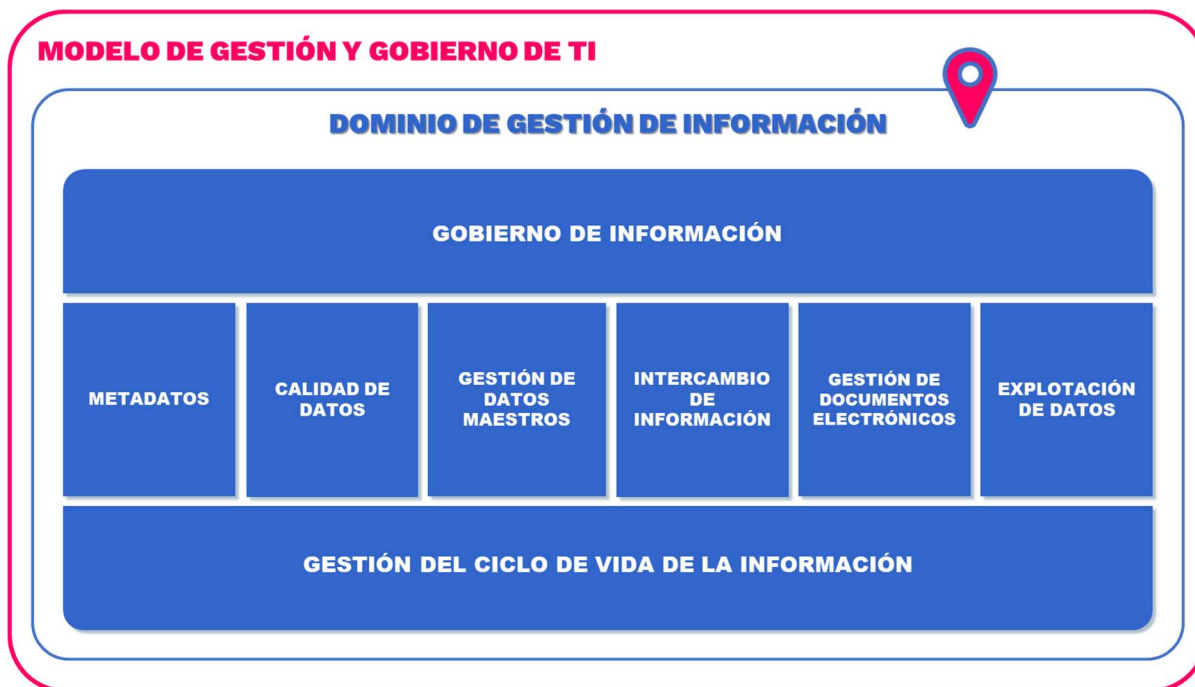


Ilustración 3. Ubicación de la Guía de Gestión de Información en el contexto del Modelo de Gestión y Gobierno de TI

1.2. Propósito de esta guía

General

Orientar la gestión de la información en las entidades del Estado colombiano.

Específicos

- Brindar orientación de cómo desarrollar ejercicios de calidad de datos.
- Ofrecer una guía para gestionar los Metadatos y los datos maestros.

1.3. A quién va dirigida

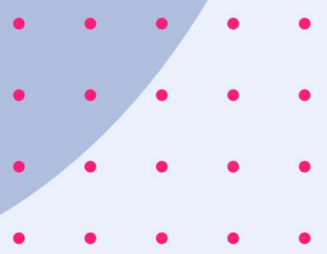
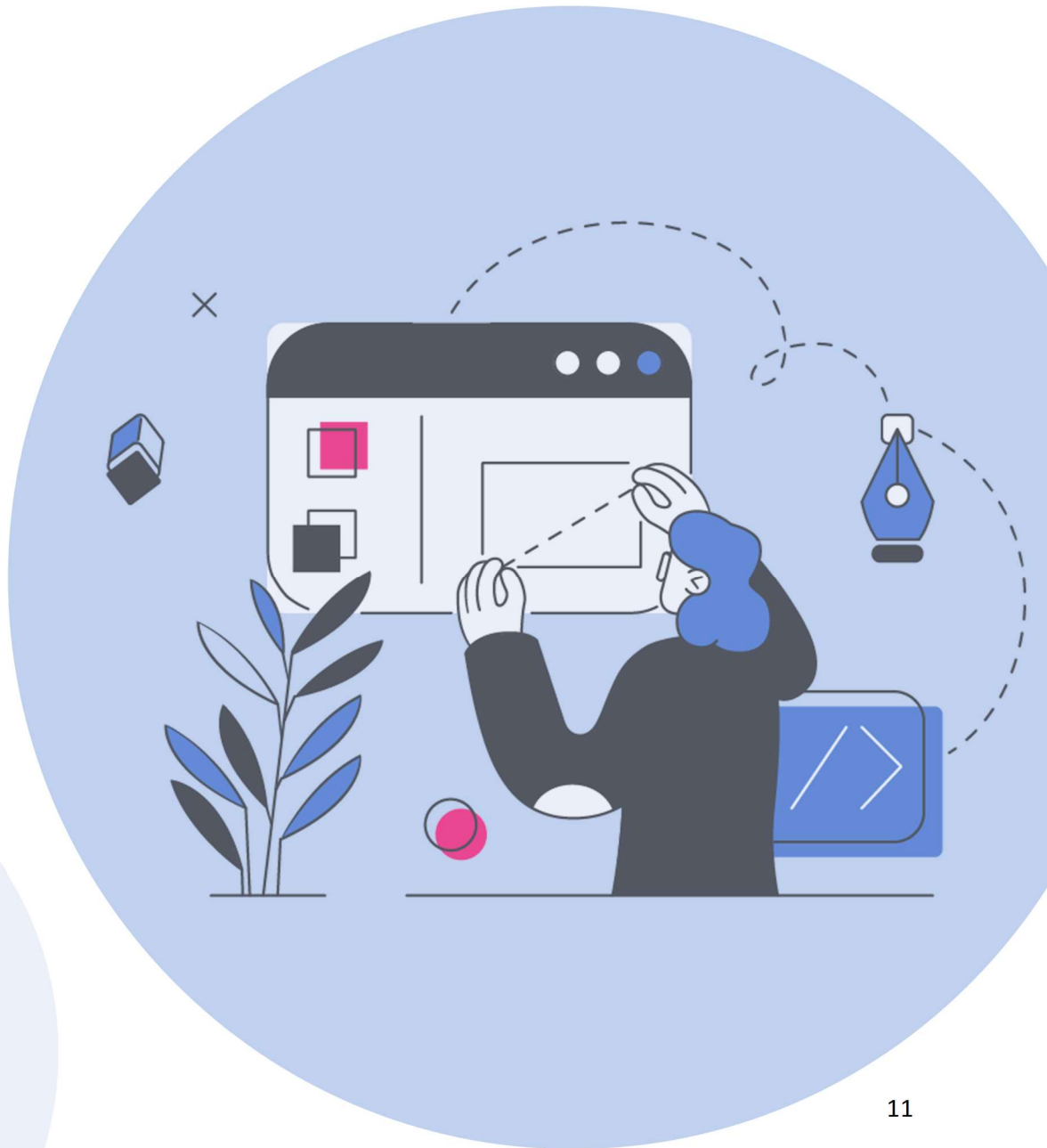
Esta guía va dirigida a las áreas de tecnología, planeación, gestión de información y roles que gestionan datos e información en las entidades.:



Ilustración 4. Audiencia

2.

Lineamientos

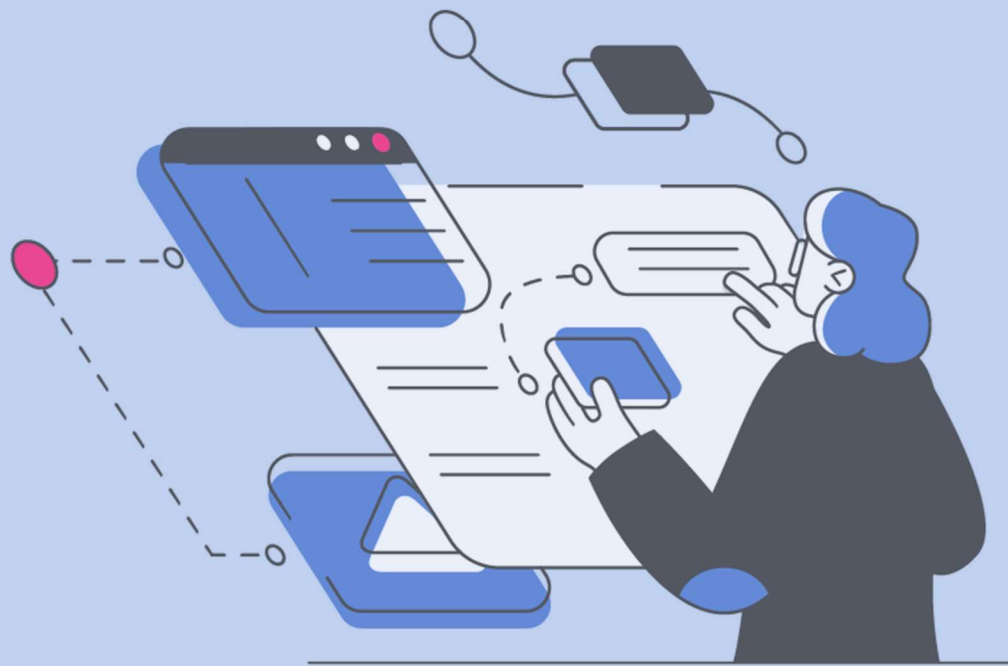


Código	Nombre	Descripción
MGGTI.LI.GI.01	Gobierno de la información	Las entidades de la administración pública deben definir un modelo de gobierno que gestione las políticas, responsabilidades, decisiones y métricas para ejercer autoridad sobre la información.
MGGTI.LI.GI.02	Gestión de la calidad de los datos	Las entidades de la administración pública deben definir y desarrollar una estrategia para diagnosticar, medir, monitorear y establecer acciones que permitan contar con información de calidad para la toma de decisiones.
MGGTI.LI.GI.03	Gestión de documentos electrónicos	Las entidades de la administración pública deben establecer un programa para la gestión de documentos y expedientes electrónicos.
MGGTI.LI.GI.04	Marco de Referencia Geoespacial	Las entidades de la administración pública deben adoptar las directrices y lineamientos encaminados a facilitar los procesos de gestión geoespacial, de acuerdo con lo definido en el Marco de Referencia Geoespacial de la ICDE. ¹
MGGTI.LI.GI.05	Publicación de los servicios de intercambio de información	Las entidades de la administración pública deben exponer sus servicios de intercambio de información a través de la Plataforma de Interoperabilidad del Estado colombiano.
MGGTI.LI.GI.06	Acuerdos de Nivel de servicio para intercambio de información	Las entidades de la administración pública deben establecer Acuerdos de Nivel de Servicio (ANS) que permitan el intercambio de información de calidad entre sus dependencias o con otras instituciones.
MGGTI.LI.GI.07	Uso del código postal colombiano	Las entidades de la administración pública deben en el diseño de sus componentes de información identificar aquellos a los que se les deba aplicar el código postal.
MGGTI.LI.GI.08	Explotación de datos	Las entidades de la administración pública deben seleccionar técnicas y herramientas que faciliten el análisis de los datos, habilitando la toma de decisiones en base a información de calidad.

Tabla 1. Lineamientos del Dominio de Gestión de Información

¹ <https://www.icde.gov.co/marcos/marco-de-referencia-geoespacial>

3. Etapas



A continuación, se describen las etapas y actividades que deben desarrollarse o ejecutarse como parte de la gestión de la información; en este caso, las etapas descritas no obedecen a una serie de pasos con un orden lógico, puesto que todas las etapas se van desarrollando de forma paralela, como parte de la gestión de la entidad.

3.1. Gestión del Ciclo de vida de los datos

La información es materia prima que recorre las instituciones en todos los sentidos, como cualquier otro activo tiene un ciclo de vida que debe ser gestionado. Entender el ciclo de vida de los datos, es clave para que las entidades la conviertan en un activo que genera valor; es por esto, el dominio de gestión de información define una guía conceptual que facilite el entendimiento del ciclo de vida de los datos y resalte su importancia para las instituciones. para sumergirse en la explicación, consultar la Guía Conceptual del Ciclo de Vida de los datos, que hace parte del Dominio de gestión de información y por ende complementa esta guía.

3.2. Gobierno de datos e Información

“La información es un asunto de todos”, esta frase que debe ser un principio orientador de las arquitecturas de datos debería quedar en la mente de todos los funcionarios, y colaboradores de las entidades del Estado colombiano. Todas las dependencias de la entidad deben estar involucradas con el Gobierno de la Información; no es responsabilidad exclusiva de las oficinas de Tecnologías o sistemas de las entidades, de hecho, no son las principales responsables.

Existen diferentes modelos o arquetipos que permiten definir la estructura de Gobierno de la Información a nivel institucional, cada institución de acuerdo con sus necesidades escogerá el más pertinente, se recomienda que el Gobierno de la información sea liderado por las áreas misionales o estratégicas de la entidad, apoyándose en las oficinas de tecnologías o sistemas de información.

Para consultar en detalle cómo definir y gestionar el Gobierno de la Información, consultar la Guía Específica de Gobierno de Información (que hace parte de este dominio).

3.3. Calidad de datos

Disponer de datos con baja calidad implica entre otras: pérdida de información, reprocesos, ineficiencia operativa, conflictos entre dependencias, mala toma de decisiones, insatisfacción de la ciudadanía; costos de oportunidad asociados a dejar de hacer análisis más precisos o elaborados, etc.

Contar con datos de calidad no viene dado sólo por establecer un proceso o realizar unas acciones de diagnóstico específicas, para lograr calidad de los datos y la información es importante tener claro que realizar una adecuada gestión de la información en sí misma, dará lugar a mejoras sustanciales de la calidad, por ejemplo, si una entidad

realiza ejercicios de calidad de datos pero no cuenta con un Gobierno de Información definido o no gestiona sus Metadatos, los resultados que obtendrán no serán los esperados y en todo caso serán temporales.

En la “Guía Específica de Gestión de Calidad de Datos”, se explica en detalle cómo definir y desarrollar una estrategia de calidad de datos, teniendo en cuenta dimensiones, datos críticos, definición de reglas mecanismos de medición, etc.

3.4. Identificación y gestión de Datos maestros

Todas las entidades cuentan con datos que son transversales a todas las dependencias, estos datos deben tener un mismo significado y valor para cualquiera de las dependencias; sin embargo, los silos y el desgobierno de la información, provocan que los datos no sean gestionados adecuadamente y que se presenten problemas de calidad en estos.

Existen cuatro grandes tipos de datos: transaccionales, maestros, de referencia y abiertos (algunas veces se suele incluir los metadatos como un quinto tipo).

Dato Maestro: Conjunto de datos central, esencial y transversal en una organización definido y establecido como única fuente de verdad. Puede ser compartido por diferentes sistemas de información de la organización y en ocasiones por fuera de la misma.

Dato transaccional: Son los datos generados o capturados por una organización como resultado de la ejecución de una operación en un sistema o entre sistemas y que pueden ser recopilados o almacenados.

Datos de referencia: Conjunto de datos proveniente de estándares internos o externos que permite la clasificación, la caracterización y la categorización de datos en la organización. Generalmente hace parte de los datos maestros. Ejemplo datos de la DIVIPOLA, los códigos CIU.

Datos abiertos: son información pública dispuesta en formatos que permiten su uso y reutilización bajo licencia abierta y sin restricciones legales para su aprovechamiento².

La gestión de los datos maestros incluye: la gestión de los datos maestros en sí mismos y la gestión de los datos de referencia; su gestión es de suma importancia para mejorar la calidad de la información y para que las entidades y el Estado cuenten con información de valor para tomar decisiones.

En la “Guía Específica de Gestión de Datos Maestros”, se explica en detalle como las entidades pueden estructurar su estrategia de gestión de datos maestros.

² Concepto tomado del Plan Nacional de Infraestructura de Datos.

3.5. Identificación y Gestión de Metadatos

Los metadatos son históricamente, una de las áreas de gestión de datos a la que menos relevancia se le da; sin embargo, es supremamente importante y es difícil pensar en una adecuada gestión de la información, sin una gestión de metadatos.

Un ejemplo típico para explicar en términos prácticos qué es un metadato, es hacer el símil de instituciones vistas como una biblioteca. Cuando se va a una biblioteca en busca de un libro, el sistema permite hacer búsquedas por: autor, libro, ISBN, tipo de obra, país, etc., Una vez se encuentra en el sistema el libro deseado, un código que identifica el libro permite encontrarlo rápidamente entre los estantes. Para hacer esto posible, en las bibliotecas cada libro cuenta con una ficha que describe toda la información que permite caracterizar el libro (autor, libro, ISBN, tipo de obra, país, etc.) y codificarlo; Sin esta ficha sería supremamente difícil encontrar un libro, entre más grande la biblioteca más pequeña es la posibilidad de encontrar el libro rápidamente. Si una institución fuera una biblioteca, para este ejemplo se puede establecer una analogía entre los libros y los datos; los datos que hacen parte de la ficha del libro corresponden a los metadatos.

La siguiente ilustración, muestra el proceso que debe seguir una entidad para realizar una gestión de metadatos madura:

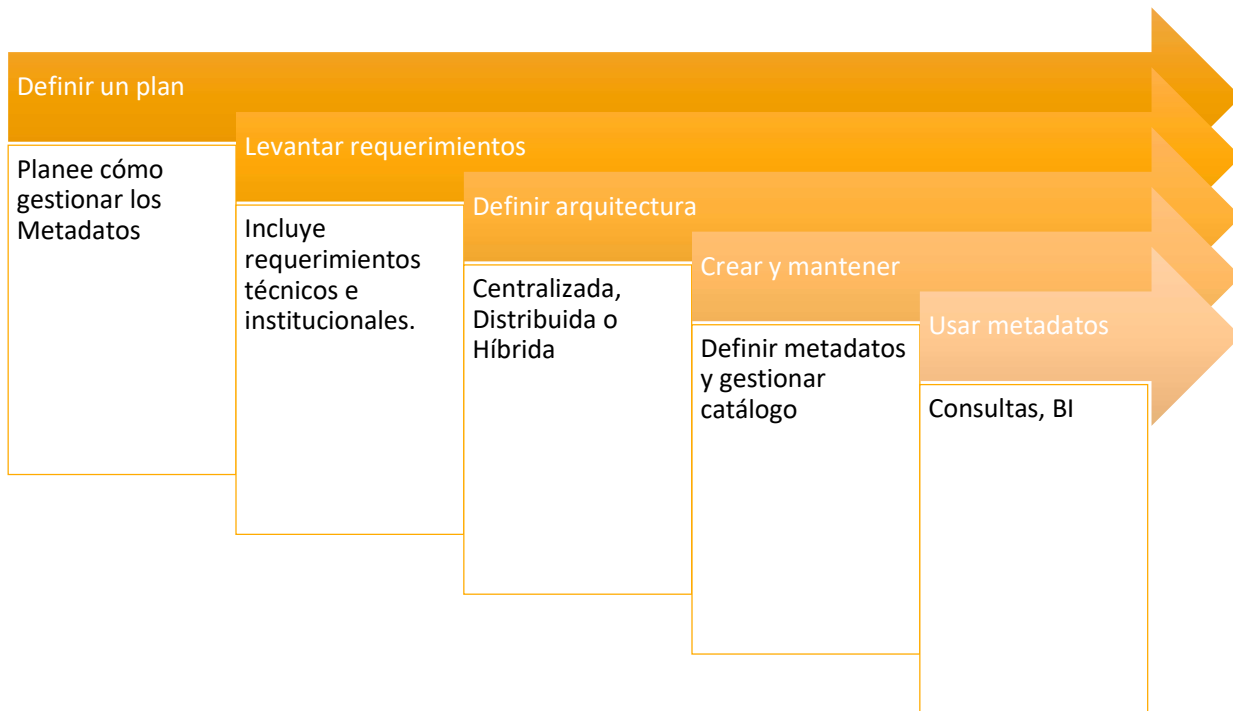


Ilustración 5. Proceso para gestionar Metadatos

Catálogo de Metadatos

Existen soluciones tecnológicas que permiten la gestión avanzada de metadatos; sin embargo, como la mayoría de las entidades no pueden acceder a una herramienta de este tipo, se propone realizar la gestión de Metadatos, a través de un catálogo:

Id Entidad	Entidad	Id Atributo	Atributo	Único	Requerido	Repetible	Tipo	Longitud	Valores válidos	Valor por defecto	Responsable	Definición
	Contrato	C01	Modalidad de contrato	No	Sí	Sí	Cadena	100	Contratación directa, Contratación pública, Contratación privada, Otras modalidades	NA	Área de contratación	Describe la forma de contrato por la que se seleccionó al contratista.

Ilustración 6. Catálogo de Metadatos

Como archivo anexo se puede consultar el Catálogo de Metadatos, el archivo se encuentra caracterizado.

Sí tiene documentado el Modelo Físico de Datos (modelo entidad relación u otro modelos de datos), es posible que este catálogo puede estar descrito como parte del Modelo a través del diccionario de datos de ellos modelos de datos de los sistemas que posee la entidad o como parte de las fichas de diagnóstico de calidad de los datos; en caso de ser así, se debe definir a través del Gobierno de la información, en cual artefacto o herramienta se puede consultar la caracterización de los Metadatos; Se debe evitar duplicar información.

3.6. Explotación de los datos

El fenómeno que se conoce como la “Cuarta Revolución Industrial”, ha permitido que técnicas para el análisis y la explotación de datos que habían sido definidas el siglo pasado, pueden ser utilizadas de manera eficiente debido a la gran capacidad de procesamiento de datos que se genera por la explosión de estas nuevas tecnologías. El CONPES 3920 de 2018, define la Política Nacional de Explotación de Datos y en consonancia con esta política las entidades del Estado deben iniciar o mejorar la explotación de sus datos para tener información de valor y conocimiento. La explotación de datos no se genera al mismo nivel en todas las entidades debido a su tamaño y necesidades; sin embargo, es importante que todas las entidades del Estado colombiano se concienticen de la importancia de explotar sus datos. Algunas entidades usarán tecnologías como Big Data, Machine Learning o Blockchain; sin embargo, a otras les bastará con explotar datos a través de analítica tradicional, como por ejemplo la inteligencia de negocios - BI³.

Para orientar a las entidades en el desarrollo de ejercicios de explotación de datos, de acuerdo con su nivel de madurez y teniendo en cuenta el uso de tradicionales y nuevas tecnologías, se plantea la “Guía Conceptual de Explotación de Datos”, la cual se puede consultar para tener mayor detalle de los temas y recomendaciones relacionados con la explotación de datos.

Cuando una entidad decide emprender una iniciativa de explotación de datos, se recomienda que sea definida desde la arquitectura empresarial y que la arquitectura de información objetivo incluya los componentes de arquitectura que habiliten la explotación de datos en la entidad. Para más información consultar la “Guía General

³ De sus siglas en inglés: Business Intelligence.

del Dominio de Arquitectura de Información” y la “Guía Específica de Componentes de Arquitectura de Información”, que hace parte del Modelo de Arquitectura Empresarial.

3.7. Apertura de datos

Las entidades del Estado colombiano deben identificar sus conjuntos de datos abiertos y diseñar mecanismos que permitan generarlos, automatizarlos y publicarlos en el Portal de Datos Abiertos del Estado o del Distrito según corresponda. Si los datos abiertos son publicados en el portal de datos de Bogotá se debe validar que los datos sean federados al portal de Datos.gov.co

Las entidades deben gestionar el ciclo para el proceso de apertura, mejora y uso de los datos; este ciclo se recomienda que se defina para un periodo de dos años. Sus principales fases son:

- Establecer un plan de apertura de datos.
- Estructurar y publicar los datos.
- Comunicar y promover el uso de los datos.
- Monitorear la calidad y el uso de los datos abiertos.

Los datos abiertos publicados en el portal que para tal fin tiene el Estado colombiano, pueden y deben ser consultado por los ciudadanos, la academia, empresas privadas, periodistas y/u otros actores que puedan a través de esa información generar valor o fomentar la transparencia y el control social.

Para consultar en detalle la documentación, herramientas y datos publicados; consultar en <https://datos.gov.co>.

3.8. Gestión de Documentos electrónicos

Independientemente de si la información es un documento o un contenido, la concepción es la misma si es digital o “electrónico”. De hecho, los documentos en papel (u otros soportes analógicos) se digitalizan. También se gestionan correos electrónicos, documentos generados desde herramientas tecnológicas como un CRM o un ERP y en diversos formatos como XML, CSV, HTML o documentos generados a través de procesadores de texto y formularios.

Partiendo de la premisa de que para implementar la gestión de documentos electrónicos se requiere de tecnología con funcionalidades particulares que cubran las actividades necesarias para la gestión durante su ciclo de vida, se hace necesario que cada organización antes de adquirir herramientas tecnológicas realice un análisis de sus necesidades y una definición de sus requerimientos previos.

Los Administradores de Contenidos Empresariales o herramientas ECM permiten la gestión de cualquier tipo de activo digital en todas las etapas de su ciclo de vida desde su creación hasta su disposición final, a través de automatización de flujos de trabajo, gestión de reglas, revisión compartida, generación de comentarios, clasificación de la información y gestión de etiquetas entre otros. Este proceso requiere de una etapa de planeación y alistamiento de información y puede luego involucrar el uso de tecnologías para el reconocimiento

y clasificación de los documentos, herramientas de extracción de datos de forma automática y el almacenamiento de metadatos.

Un sistema de gestión documental contempla la definición e implementación de principios, lineamientos y estándares para la gestión del ciclo de vida de la información (creación, uso, mantenimiento, acceso, disposición final, conservación y preservación) contenida en cualquier soporte y medio.

La información es un activo importante para toda organización. Facilita la toma de decisiones, asegura la transparencia de la gestión y constituye su memoria institucional. Este componente es esencial y por ello se tiene en cuenta dentro del Dominio de Gestión de Información.

El Archivo General de la Nación ha dispuesto una serie de guías para orientar a las entidades en temas relacionados con la gestión de documentos electrónicos, entre estas se destaca: la Guía de Implementación de un Sistema de Gestión de Documentos Electrónicos de Archivo⁴ y también existe la guía del Marco de Referencia: Guía para la gestión de documentos y expedientes electrónicos.

3.9. Identificación y gestión de datos geoespaciales

Las entidades del Estado colombiano que como parte del desarrollo de su misión deban gestionar datos geoespaciales (geográficos), deben cumplir con los lineamientos descritos en el Marco de Referencia Geoespacial definido por la Infraestructura de Datos Espaciales – ICDE. Este marco *es un instrumento que proporciona las directrices y lineamientos encaminados a facilitar y optimizar los procesos de gestión geoespacial. Esta gestión incluye los procesos y/o actividades requeridas en la planeación, producción, armonización, mantenimiento, actualización, disposición y reutilización de recursos geoespaciales del territorio colombiano* (ICDE, 2021).

Entre las principales actividades relacionadas con la gestión de la información, las entidades del Estado colombiano deben elaborar el Inventario de Datos Geoespaciales producidos, así como definir y disponer los Datos Geográficos Fundamentales de Colombia a través de la plataforma tecnológica de la ICDE; los *Datos Fundamentales deben presentarse en formato digital interoperable que permita el intercambio de manera abierta bajo estándares internacionales* (ICDE, 2021) todo esto, de acuerdo con los lineamientos definidos por la ICDE a través del Marco de Referencia Geoespacial.

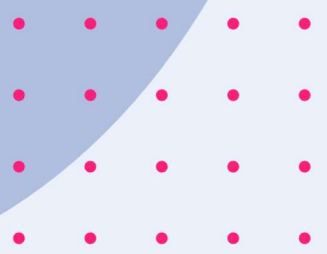
Se puede entender que los datos fundamentales a los que hace referencia el Marco de Referencia Geoespacial pueden ser considerados datos maestros o datos referencia.

Además de buscar el fortalecimiento del intercambio de información, el Marco de Referencia Geoespacial busca que se implementen procesos de análisis y explotación de la información usando Big Data, Inteligencia Artificial y Machine Learning.

⁴ https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicaciones/ImplementacionSGDEA.pdf

Para consultar en detalle las obligaciones y los instrumentos definidos dentro del Marco, consultar la información oficial en <https://www.icde.gov.co/marcos/marco-de-referencia-geoespacial>.

4. Roles



A continuación, se listan los roles que deberían estar presentes durante la gestión del ciclo de vida de los datos.

Rol	Responsabilidades
Arquitecto empresarial	<ul style="list-style-type: none"> • Alinear de la gestión de la información y la estrategia de TI con las metas de la entidad. • Identificar riesgos en la gestión de la información que impacten la arquitectura empresarial y el desarrollo de la estrategia institucional.
Arquitecto de información	<ul style="list-style-type: none"> • Definir y gestionar la arquitectura de datos y metadatos. • Proponer mecanismos adecuados para la explotación de datos. • Define lineamientos y estándares para la gestión de información • Tener en cuenta las definiciones de otros Marcos del Estado colombiano que afecten la gestión de la información y velar por su cumplimiento. • Diseñar y ejecutar procesos de integración y consolidación de datos
Analista de calidad de datos	<ul style="list-style-type: none"> • Determinar la idoneidad de los datos para su uso. • Supervisar la condición de los datos. • Analizar la causa raíz de los problemas de los datos. • Identificar y organizar procesos y mejoras técnicas que contribuyan a mejorar la calidad de los datos.
Administrador de datos	<p>Las funciones de este rol están definidas en el decreto 1389 de 2022 del Modelo de gobernanza de la infraestructura de datos, las cuales se listan a continuación:</p> <ul style="list-style-type: none"> • Diseñar, asesorar, impulsar y poner en marcha las acciones y actividades necesarias para implementar las recomendaciones hechas por el Comité Nacional de Datos, en el marco del Plan Nacional de Infraestructura de Datos. • Liderar la planeación, coordinación e implementación de estrategias en la entidad para la debida gestión de los datos durante todo su ciclo de vida. • Liderar la definición e implementación del modelo de gobierno de datos de la entidad, y coordinar las acciones en el nivel táctico y operativo necesarias para su adecuada implementación. • Elaborar y mantener actualizado el inventario de datos y el diccionario de datos de la entidad, donde se identifique dentro del ciclo de vida de los datos, a aquellos funcionarios que estén directamente relacionadas con la creación, procesamiento, almacenamiento, intercambio, uso y análisis, archivo y preservación de los datos en todas las áreas y niveles del gobierno. • Articular esfuerzos con el director, jefe de oficina o coordinador de Tecnologías y Sistemas de la Información de la entidad de la entidad sobre asuntos referentes a la Política Digital que se relacionen con el ciclo de vida de los datos. • Coordinar con el director, jefe de oficina o coordinador de Tecnologías y Sistemas de la Información de la entidad la implementación de lineamientos del Marco de Referencia de Arquitectura Empresarial MRAE, asociados con los datos y su gestión.

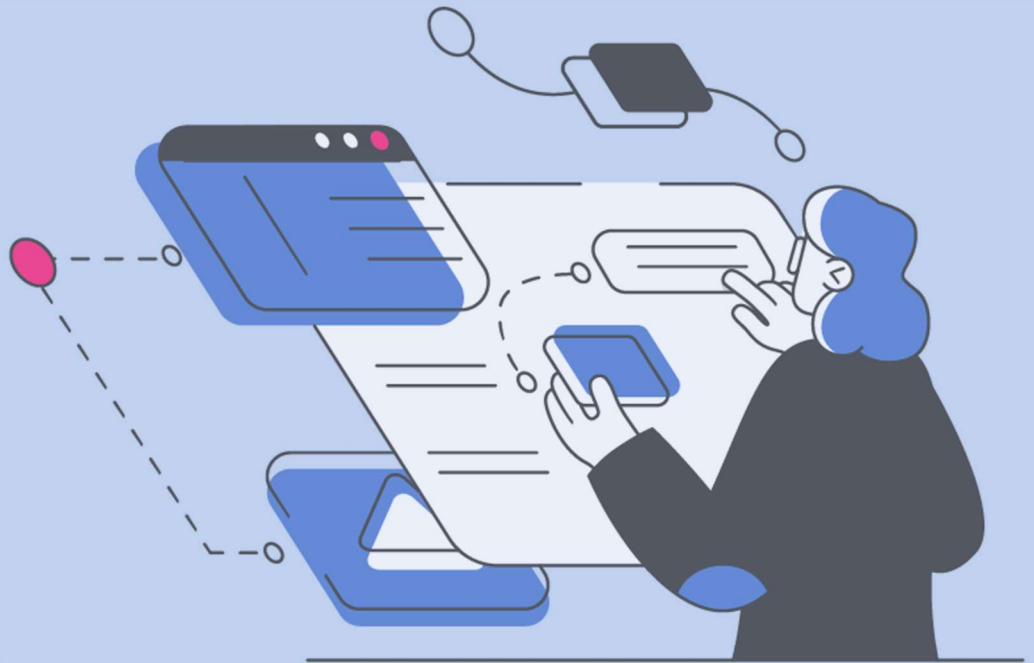
	<ul style="list-style-type: none"> Realizar el seguimiento de la evolución de las capacidades y competencias con relación al uso y explotación de datos en el interior de la entidad.
Dueño de datos	<ul style="list-style-type: none"> Garantizar un uso adecuado de los datos. Realizar la definición conceptual de la información. Establecer políticas para mantener y manejar los datos una vez almacenados. Responder por la disponibilidad de los datos.
Administrador de bases y repositorios de datos	<ul style="list-style-type: none"> Administrar y gestionar las bases y repositorios de datos Realizar afinamiento y optimización de las bases y repositorios. Mantener documentados los modelos de datos y objetos (triggers, procedimientos almacenados, índices, vistas, otros) de las bases de datos. Realizar copias de seguridad de las bases de datos y repositorios. Realizar restauración y/o recuperación de copias de seguridad Demás actividades propias de la administración de bases de datos.
Oficial de protección de datos personales	<ul style="list-style-type: none"> Define, estructura y orienta la implementación aplicación de la política de protección de datos personales
Analista o científico de datos	<ul style="list-style-type: none"> Realizar visualización y análisis de datos. Construir y diseñar modelos de analítica de datos.

Tabla 2. Roles de la Arquitectura de Información



(Tenga en cuenta) Una persona puede desempeñar uno o más roles.

5. Caso práctico



Nota: El caso que se explica a continuación describe entidades y eventos ficticios; la solución planteada al problema no necesariamente refleja lo que una entidad con funciones similares debiese decidir; la situación busca facilitar el entendimiento de las decisiones que deben tomarse como parte de la definición de la arquitectura de información.

5.1. Contexto

En el sector educación las instituciones educativas son una persona jurídica. Desde el punto de vista administrativo del sector tienen un *identificador* designado por el nivel nacional. Cada institución educativa cuenta con una o más *sedes*. Cada sede tiene un identificador de sede. Las sedes dependen administrativamente de la institución educativa.

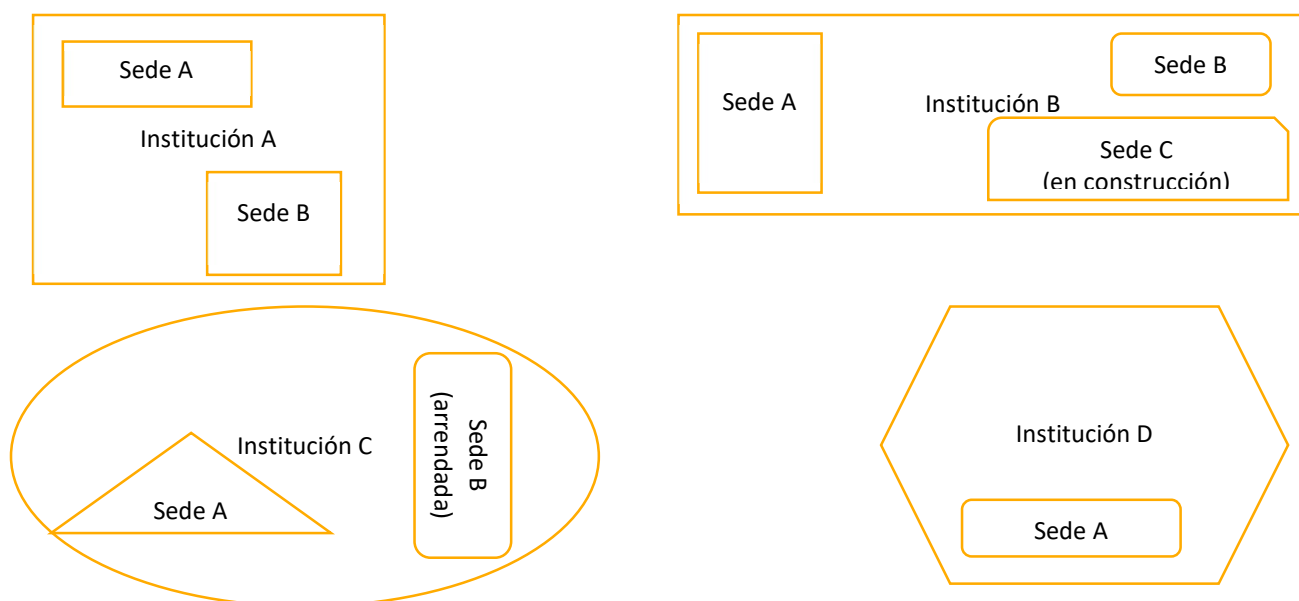


Ilustración 7. Esquema de la estructura de sedes dependientes de instituciones educativas

El área de contratación puede arrendar infraestructura para prestar el servicio educativo. (En el diagrama, la Institución educativa C tiene la sede B arrendada, tal vez porque está cerrada temporalmente la propia por cuenta de mejoras locativas). La infraestructura arrendada, que se constituye en una sede, dicha área la identifica por medio del *número del contrato*.

El área de construcciones escolares realiza mantenimiento a la infraestructura escolar y construye nuevas instalaciones (En el diagrama, la Institución educativa B cuenta con dos sedes en operación y pronto contará con una nueva sede, la cual está en construcción). Dentro de una sede puede haber varias edificaciones, construidas en diferentes épocas, por lo cual tienen necesidades de mantenimiento diferentes. Identifica cada edificación de manera independiente con un número designado por la propia área de construcciones.

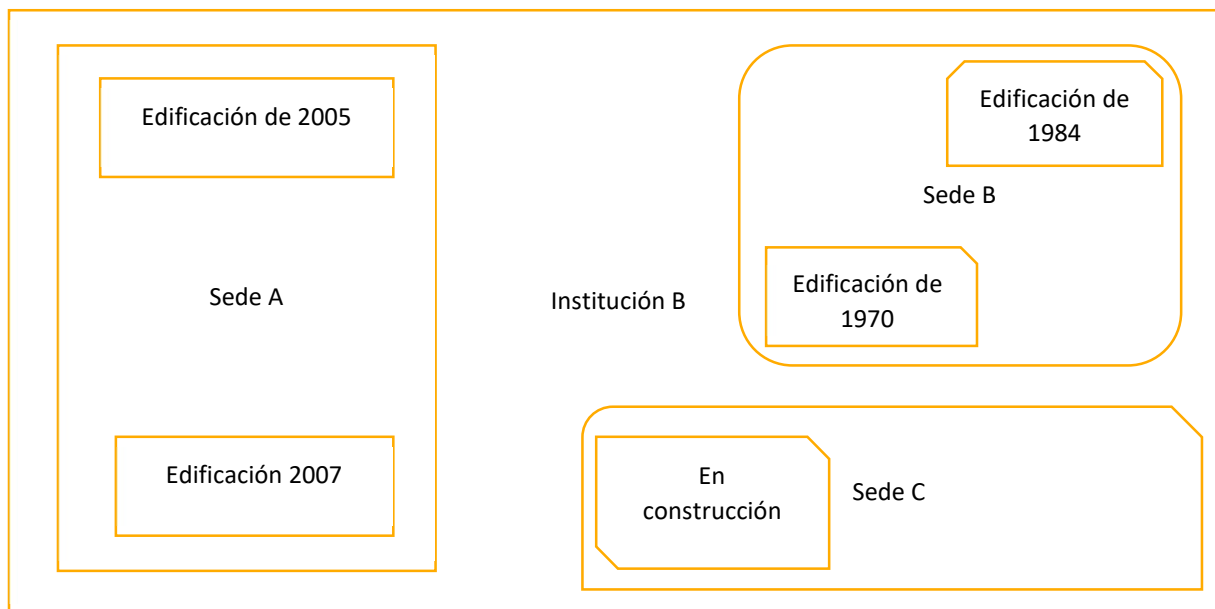


Ilustración 8. Esquema de la estructura de edificaciones dependiendo de sedes educativas

El repositorio de documentación de la entidad territorial tenía un viejo diccionario de datos (metadatos⁵) de una tabla denominada *directorío de establecimientos educativos*.

Los datos del directorio de establecimientos educativos están albergados en una base de datos relacional que mantiene integridad referencial con la base de datos de recursos humanos, lo cual permite conocer la ubicación de docentes y administrativos que atienden el servicio. Pero la base de datos de contratos y la de mantenimiento de edificaciones son bases de datos aisladas que se administran por medio de una hoja de cálculo. No tienen metadatos que expliquen qué atributos manejan ni cómo entienden el nombre y contenido de cada atributo.

5.2. Situación

El área jurídica deseaba contestar una denuncia de detrimento patrimonial en la que se ve involucrada una institución educativa en la que informan que el área de construcciones está haciendo mejoras. No tiene sentido para los abogados que una mejora se constituya en detrimento patrimonial. El caso es que le tomó al área jurídica un mes darse cuenta de que el problema radicaba en que la sede en la que se estaban realizando mejoras era arrendada. Ni el rector de la institución educativa recordaba, ni el área de construcciones sabía, que dicha sede era arrendada y habían dado viabilidad al proyecto hacía varios meses.

El tema fue puesto en la palestra pública por parte de la prensa. Fue necesario dar las explicaciones a la ciudadanía, costó mucho trabajo aclarar la situación ante los entes de control, enderezar las cosas para que el dueño del predio asumiera los costos en que se había incurrido y anular el contrato de obra para no continuar realizando las adecuaciones a costa del erario. Se hizo prioridad para la entidad ajustar las bases de datos para que ese tipo de errores no volvieran a ocurrir. Parecía que se trataba de un problema de falta de conectividad

⁵ Ver Guía específica de Metadatos.

de la tabla maestra⁶ de contratos dando contexto a los registros administrativos del área de construcciones escolares y a la tabla de referencia de instituciones educativas y sedes.

5.3. Primeros pasos

El subsecretario académico sabía que un arreglo coyuntural no evitaría que volviese a ocurrir. Ordenó una solución de fondo. El impulsor del esfuerzo sería evitar los costos asociados a este tipo de errores.

Para resolver la problemática planteada la entidad asignó un Comité ad hoc: un profesional de planeación, área que administra el directorio de establecimientos y sedes educativas, un profesional de desarrollo institucional para que ajustara los procedimientos que fuesen necesarios y un administrador de datos dedicado a apoyar la identificación y definición de las entidades lógicas de información dentro del Modelo de Información Institucional.

Siguiendo las Guías del Dominio de Gestión de la Información, recorrieron algunos pasos:

Paso 1: Definición de las dimensiones de la calidad de los datos.

El problema estaba concentrado en un tema muy puntual: el directorio de establecimientos. El profesional de planeación explicó que el ente rector del directorio de establecimientos a nivel nacional era el Ministerio de Educación y que se habían establecido reglas, entre ellas, que se creaba, actualizaba o eliminaba un establecimiento educativo respaldado en una resolución dada por la entidad territorial. Así que acordaron que las dimensiones que les interesarían eran: exactitud y completitud acorde con lo que ya se había definido en el área de planeación⁷. La exactitud se definiría en relación con las resoluciones de creación de las instituciones educativas y sus sedes. La completitud se definiría respecto a lo alojado en el directorio nacional de establecimientos educativos que administraba el Ministerio de Educación.

Paso 2: Definición de la estrategia

1. Se tomaría como base de la situación actual lo pertinente que hubiera en el repositorio institucional respecto a la Arquitectura Empresarial realizada hacía un par de años.
2. Se le solicitó a un analista de calidad de datos que midiera la calidad de las bases de datos de Contratación y de Construcciones escolares.
3. Y se asignó a otro analista de calidad de datos, con experiencia como integrador, para unificar las tres bases de datos. Como la información de las áreas de contratos y construcciones escolares eran hojas electrónicas, debió realizarlo de manera manual.

El equipo especificó la razón de ser del directorio de establecimientos educativos y sedes. ¿Para qué el directorio? La identificación de las sedes educativas permite el cruce de la información que recopilan las áreas de la entidad y el intercambio de información con otras entidades. Los datos asociados permiten realizar agrupaciones de datos (por ejemplo, número de estudiantes, de docentes, etc.) o clasificaciones (por ejemplo, colegio oficial o no oficial) con el objeto de informar decisiones. La dirección es útil desde el punto de vista de la gestión (proveeduría, procesos de mantenimiento, etc). Las coordenadas son útiles para el análisis geográfico y el cruce con otros datos de carácter geográfico o espacial⁸.

⁶ Ver Guía específica de Datos maestros.

⁷ Ver Guía específica de Calidad de datos.

⁸ Ver Guía específica de Explotación de datos.

El analista de calidad entregó un informe al Comité ad hoc. La medición de las dimensiones de exactitud y completitud fueron insatisfactorias. El analista de calidad hizo la observación de que las dos bases de datos en hojas electrónicas carecían del código único de identificación del establecimiento educativo y de sede por tanto no era fácil cruzar información y había establecimientos duplicados. Las áreas las identificaban por medio de números asignados dentro de la propia área o por el nombre. Intentó cruzarlas por el nombre, pero éste no necesariamente coincidía con el del directorio o entre las otras dos hojas electrónicas. Intentó verificar que fuera la misma institución por medio de la dirección, pero ésta no necesariamente coincidía ni en la forma de escribirla, ni en la nomenclatura.

El Comité ad hoc se dio cuenta que trascendía el tema de datos de referencia e involucraba un problema relacionado con Datos maestros y con Calidad de los datos.

5.4. Pasos adicionales

Paso 3: Identificación de los datos críticos y de las reglas de negocio.

Realizaron una reunión con los dueños de los datos. De la conversación obtuvieron información reveladora:

- a. El directorio de sedes tenía registrada la dirección de la entrada al área administrativa de los colegios. El área de construcciones mantenía la dirección asociada al recibo del agua y el área de contratación la que figuraba en el título de propiedad del inmueble. El problema podía ser de nomenclaturas antiguas o el hecho de que las sedes de las instituciones educativas suelen ocupar una manzana completa y cuentan con entradas por varios costados.
- b. El directorio de sedes, desde el punto de vista de la prestación del servicio, sólo contenía los datos de las sedes en operación, pero el directorio del área de construcciones incluía las sedes en proceso de construcción o las que estaban fuera de servicio temporalmente por reforzamiento estructural u otros temas (intervenciones que suelen tomar poco más de un año y obligan al cierre de la sede transitoriamente). Y el área de planeación informó que mantenían un cuadro de referencia con la identificación, ubicación y caracterización de las instituciones y/o sedes antiguas que ya no están en operación, pero que les era importante para realizar análisis de evolución de la matrícula por zona geográfica a lo largo de los años.

Era necesario actualizar el Modelo de Información Institucional⁹. Se reunieron con el equipo de modelado de información institucional y actualizaron la capa conceptual y la lógica, dejando como proyecto posterior, una vez visualizado todo, actualizar la capa física y generar requerimientos a la arquitectura empresarial en el dominio de arquitectura institucional y arquitectura de datos. Actualizaron el concepto de *contrato* para especificar mejor los contratos y poder identificar los que eran sedes para el servicio educativo, y añadieron el de *construcción escolar* en referencia a las sedes en construcción. También especificaron mejor el concepto de *dirección* en cuanto que la hay “resolutiva”, es decir, la definida en la resolución de creación, “administrativa”, “de carga y descarga” y “de entrada de estudiantes”.

Paralelamente trabajaron en especificar los datos críticos y las reglas de negocio.

El directorio de establecimientos educativos y de sedes eran datos críticos en cuanto que manejaban datos transversales a todas las áreas dando contexto a los registros administrativos, transaccionales y a los análisis que

⁹ Ver Guía específica del Modelo de Información Institucional.

realizaba la institución, pero también eran utilizados para caracterizar o clasificar otros datos, o para relacionar los datos con información externa a la entidad. Eran dualmente datos maestros y de referencia.

Como regla, se determinó evitar que la nomenclatura de escritura del atributo “dirección” fuera diferente en cada conjunto de datos. Se investigó y estableció la regla a aplicar y se generó el requerimiento, además se incluyó el código postal asociado al establecimiento educativo.

Como regla, se determinó que todo conjunto de datos que manejara el atributo “establecimiento educativo” o “sede” añadiera el identificador correspondiente como medio para permitir la integridad y el intercambio de información.

Se acordó que los conjuntos de datos debían tener asociada una regla que obligase a tener un número de contrato si la sede fuese arrendada como medio de identificación de tal situación.

Paso 4. Identificación y priorización de las mejoras.

El comité informó a las directivas de la entidad cuán costoso era acometer el trabajo encomendado en tiempo y en recursos humanos dedicados al tema. Informó las medidas adoptadas y los requerimientos generados, pero también informó cómo las áreas de planeación y construcciones escolares tenían requerimientos adicionales que no estaban contemplados en el modelado de datos institucionales ni en las demás arquitecturas (arquitectura institucional, arquitectura de sistemas de información, arquitectura de seguridad). El área de planeación identificó debilidades en el registro de sedes cerradas, divididas o fusionadas en el tiempo para los análisis de atención del servicio educativo misional en el tiempo; y el área de construcciones escolares reconoció la necesidad de contar con una ficha de las intervenciones realizadas a cada edificación y luego poder agregar dicha información para generar reportes y análisis por institución educativa. La dirección aceptó con beneplácito lo realizado y decidió dejar los requerimientos de las dos áreas para una iteración posterior.

Paso 5. Definición de metas de mejoramiento de la calidad de los datos.

La dirección, además, solicitó que les informaran las métricas de la calidad de los datos en las hojas Excel de contratos y construcciones escolares una vez el integrador “unificara” las bases con el directorio de establecimientos y de sedes. Esperaban que sólo quedarán sin integridad referencial las sedes en construcción. De resto, todo debía quedar al 100%.

Paso 6. Desarrollo de la estrategia.

El Comité ad hoc, para poner en marcha el plan para gestionar las reglas y las normas de calidad de los datos entregaron a los encargados del Gobierno de información¹⁰ el informe con la problemática, las reglas y los requerimientos establecidos. En apoyo de estas actividades, los analistas de calidad de datos y los administradores de datos también participarían en la documentación de las reglas, en establecer formalmente los requisitos de medición y reporte de calidad de los datos, en actualizar los flujos de información¹¹ y generar metadatos¹² para las dos hojas electrónicas. Finalmente, toda la documentación y actualización de las arquitecturas fue dispuesta en el repositorio de Arquitectura empresarial.

¹⁰ Ver Guía específica de Gobierno de información.

¹¹ Ver Guía Específica de Flujos de información.

¹² Ver Guía Específica de Metadatos.

6. Artefactos



Artefacto	Tipo	Descripción
Catálogo de Metadatos	Matriz	Contiene la caracterización de los Metadatos de la institución.
Catálogo de Componentes de información	Matriz	Contiene el inventario y caracterización de los activos, flujos, servicios de información, datos abiertos y datos geoespaciales.
Planes de calidad de datos	Documento	Contiene las acciones y el plan de trabajo a ejecutar para realizar limpieza, depuración y otros a fin de obtener fuentes de datos
Planes de apertura de datos	Documento	Contiene el plan de trabajo para disponer, actualizar y promocionar los conjuntos de datos abiertos de las entidades
Modelos y algoritmos de analítica de datos	Código fuente del algoritmo	Corresponde al código fuente de los algoritmos de analítica y explotación de datos
Modelo de gobierno de datos	Documento	Contiene el esquema de gobierno definido para los datos, el cual debe incluir entre otros el esquema de toma de decisiones, políticas, indicadores,

Tabla 3. Lista de artefactos

7. Estándares y Mejores prácticas



Estándar / Mejor práctica	Descripción
DAMA-DMBOK	(DMBOK: Data Management Body of Knowledge), es un reconocido estándar en gestión de datos, desarrollado por el DAMA (Data Management Association)
ISO / IEC 11179	Estándar ISO para la gestión de Metadatos.
MIKE 2.0	Metodología de entrega de código abierto para consultores de gestión de información empresarial.
ISO/IEC 25024:2015	Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality
ISO/IEC TR 38505-2:2018	Information technology — Governance of IT — Governance of data — Part 2: Implications of ISO/IEC 38505-1 for data management
ISO/IEC 38505-1:2017	Information technology — Governance of IT — Governance of data — Part 1: Application of ISO/IEC 38500 to the governance of data
ISO/DIS 8000-51(en)	Part 51: Data governance: Exchange of data policy statements
ISO/IEC TS 38505-3:2021	Information technology — Governance of data
ISO/IEC TS 27100:2020	Information technology — Cybersecurity — Overview and concepts
ISO/IEC 20546:2019	Information technology — Big data — Overview and vocabulary
ISO 37156:2020	Smart community infrastructures — Guidelines on data exchange and sharing for smart community infrastructures
ISO/IEC 20889	Privacy enhancing data de-identification terminology and classification of techniques
ISO 27110	Marco de trabajo de Ciberseguridad
ISO 27103	Define el marco de Ciberseguridad (=NIST) para expandir SGSI
ISO 27017	Controles de Seguridad para servicios en la Nube
ISO 27006	Requisitos para las organizaciones que certifican los SGSI
ISO 37156:2020	Smart community infrastructures — Guidelines on data exchange and sharing for smart community infrastructures
ISO/IEC 30161:2020	Internet of Things (IoT) — Requirements of IoT data exchange platform for various IoT services
ISO/IEC 25010:2011	Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models
ISO/IEC 25012:2008	Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model
ISO 20614:2017	Information and documentation — Data exchange protocol for interoperability and preservation
ISO/IEC 20546	Big data — Overview and vocabulary
ISO/IEC 20547-1:2020	Big data reference architecture Part 1: Framework and application process
ISO/IEC 20547-2:2020	Big data reference architecture Part 2: Use cases and derived requirements
ISO/IEC 20547-3:2020	Big data reference architecture Part 3: Reference Architecture
ISO/IEC 20547-4:2020	Big data reference architecture Part 4: Security and privacy
ISO/IEC 23751:2022	Cloud computing and distributed platforms — Data sharing agreement (DSA) framework
ISO/IEC 19944:2017	Cloud computing — Cloud services and devices: Data flow, data categories and data use

ISO/IEC 22624	Information technology — Cloud computing — Taxonomy based data handling for cloud services
ISO/IEC 22123-1	Information technology — Cloud computing — Part 1: Vocabulary
ISO/IEC 19944-1:2020	Cloud computing and distributed platforms — Data flow, data categories and data use — Part 1: Fundamentals
ISO/IEC 19944-2	Cloud computing and distributed platforms –Data flow, data categories and data use –Part 2: Guidance on application and extensibility

Tabla 4. Estándares y mejores prácticas