



MODELO DE GESTIÓN DE LA INFORMACIÓN DE LA
UNIVERSIDAD NACIONAL DE COLOMBIA

OFICINA DE GESTIÓN DE LA INFORMACIÓN
DIRECCIÓN NACIONAL DE ESTRATEGIA DIGITAL
NIVEL NACIONAL

2023

1. Información General del Documento	
Objetivo:	El presente documento busca establecer lineamientos procedimentales y metodológicos para una efectiva gestión de la información en la Universidad Nacional de Colombia.
Alcance:	El presente Modelo de Gestión de la Información busca establecer lineamientos y metas a corto y mediano plazo en ocho áreas del conocimiento de la Gestión de la Información.
Justificación (Opcional):	<p>Como parte de las actividades de este proyecto, el Grupo de Gestión de la Información desarrolló un diagnóstico de la Política de Gestión de la Información (2021) a través del cual se identificaron problemáticas existentes en la Universidad relacionadas con el manejo de la información.</p> <p>Principalmente se encontró una débil capacidad para usar, apropiar y aprovechar la información interna y externa de la Universidad para el mejoramiento de la toma de decisiones, la transparencia institucional y la rendición de cuentas. Algunas de las causas más recurrentes que generan esta situación general son:</p> <ul style="list-style-type: none">• Debilidad en la calidad de la información.• El desarrollo de prácticas de seguimiento y evaluación que no son útiles para el mejoramiento en la toma de decisiones.• La debilidad en el gobierno estratégico y la administración de la información.• La débil orientación de la gestión y los sistemas de la información hacia la satisfacción de las necesidades de los usuarios.• El débil diseño institucional para la gestión de la información.• La debilidad de los sistemas de la información en cuanto a infraestructura y oferta de servicios
Definiciones:	Las definiciones utilizadas en el presente documento se encuentran en el apartado denominado "Glosario" al final del documento.
Documentos de Referencia (Opcional):	Los documentos de referencia del presente documento se encuentran en el apartado denominado "Referencias" al final del documento.

Condiciones Generales:	<p>Para la implementación del presente modelo es importante asignar capacidad institucional en términos de recursos financieros y recurso humano pertinente. Esto se ha venido realizando a partir de la elaboración e implementación de proyectos de inversión.</p> <p>Una cantidad significativa de objetivos y actividades incluidas en este documento pueden ser desarrolladas únicamente mediante un equipo de trabajo establecido mediante proyecto de inversión de la DNED. Sin embargo, en la medida en que se avance en madurez de la gestión de la información en la Universidad, será necesario involucrar a un mayor número de funcionarios de otras áreas y sedes.</p> <p>Adicionalmente, para una apropiada implementación del presente documento, es imperativo contar con espacios de articulación/coordinación institucional y con un esquema (al menos provisional) de gobierno de datos e información. En este sentido, el acompañamiento del Clúster de Gestión de Información, y del Comité Nacional de Estrategia Digital, es fundamental para esta tarea.</p>
-------------------------------	--

TABLA DE CONTENIDO

Introducción	14
Metodología DAMA.....	16
Contenidos del modelo	17
Dimensiones no abordadas.....	17
1. ÉTICA	19
1.1. Conceptualización.....	19
1.1.1. Relación entre los principios éticos de la Universidad Nacional y el manejo de datos	20
1.1.2. Situaciones no controlables ante la ausencia de principios éticos en el manejo de datos	21
1.1.2.1. Presentación de la información a conveniencia	22
1.1.2.2. Representación engañosa de los datos	22
1.1.2.3. Presentación de estimaciones de datos segadas	23
1.1.3. Metas	23
1.2. Propósito	23
1.2.1. Alcance Fase I	23
1.3. Desarrollo de la Fase II de ética	24
1.3.1. Proveedores	24
1.3.2. Entradas	25
1.3.3. Actividades	26
1.3.3.1. Realizar un diagnóstico de cumplimiento normativo para evaluar el estado de las prácticas actuales de la UNAL en torno al manejo ético de los datos	26
1.3.3.2. Socialización de los principios éticos para el manejo de los datos y la información	31
1.3.3.3. Formulación de metodología para la identificación de factores de riesgo éticos frente al manejo de los datos y la información.	33
1.3.4. Salidas	36
1.3.5. Usuarios	36
2. GOBIERNO DE DATOS	37
2.1. Conceptualización	37
2.1.1. Modelo de gobierno de datos	38
2.1.2. Estado de la gestión de los datos y alineación con la misionalidad de la Universidad Nacional de Colombia	40
2.1.3. Estrategia de gobierno de datos	42
2.1.4. Roles y responsabilidades en el gobierno de datos	42
2.1.5. Gestión del cambio	43
2.2. Propósito	44
2.2.1. Metas	45
2.2.2. Alcance Fase II	45
2.3. Desarrollo fase II dimensión de gobierno de datos	46
2.3.1. Proveedores	46
2.3.2. Entradas	47

2.3.3. Actividades	47
2.3.3.1. Organización del Comité de Gobierno de la Información	47
2.3.3.2. Definición de agenda de trabajo del Comité de Gobierno de la Información	48
2.3.3.3. Priorización de temas incluidos en la Política de Gestión de la Información	48
2.3.3.4. Revisión y ajuste del Modelo de Gestión de la Información (incluyendo instituciones orientadas al perfeccionamiento del gobierno de datos)	50
2.3.3.5. Formulación de Estrategia de Gobierno de la Información	50
2.3.4. Salidas	51
2.3.5. Usuarios	51
3. ARQUITECTURA DE DATOS	52
3.1. Conceptualización	52
3.1.1. Arquitectura de datos: definición	52
3.1.2. Arquitectura de datos: elementos	52
3.1.3. Arquitectura de datos: notación	53
3.1.3.1. Notación entidad-relación tradicional	53
3.1.3.2. Notación diagrama de carriles (swimlane)	53
3.1.3.3. Notación Yourdon-De Marco para diagramas de flujos de datos	54
3.2. Propósito	55
3.2.1. Metas	55
3.2.2. Alcance fase II	55
3.3. Desarrollo de la fase II de arquitectura de datos	56
3.3.1. Proveedores	57
3.3.2. Entradas	58
3.3.3. Actividades	58
3.3.3.1. Identificar las necesidades de información	58
3.3.3.2. Articular con Arquitectura Empresarial	60
3.3.3.3. Construir el diseño de arquitectura de datos	61
3.3.3.4. Construir los diagramas de flujos de datos	62
3.3.3.5. Construir la matriz CRUD (entidades/procesos)	62
3.3.4. Participantes	62
3.3.5. Salidas	63
3.3.5.1. Diseño de arquitectura de datos	63
3.3.5.2. Diagramas de flujos de datos	64
3.3.5.3. Modelo de datos empresarial	65
3.3.5.4. Levantamiento de necesidades de información	65
3.3.5.5. Matriz CRUD	65
3.3.6. Usuarios	66
4. MODELADO Y DISEÑO DE DATOS	68
4.1. Conceptualización	68
4.1.1. Tipos de modelos de datos	68

4.1.1.1.	Representaciones usadas en modelos de datos conceptuales	69
4.1.1.2.	Representaciones usadas en modelos de datos lógicos	71
4.1.1.3.	Modelo de datos físicos	72
4.2. Propósito		74
4.2.1.	Metas	74
4.2.2.	Alcance fase II	74
4.3. Desarrollo de la fase II de modelado y diseño de datos		75
4.3.1.	Proveedores	76
4.3.2.	Entradas	76
4.3.3.	Actividades	77
4.3.3.1.	Implementar metodología de identificación y documentación de los elementos para el modelamiento	77
4.3.3.1.1.	Identificar requerimientos para realizar el modelamiento	77
4.3.3.1.2.	Identificar una transacción a modelar de acuerdo con su marco normativo	77
4.3.3.1.3.	Identificar las entidades que intervienen en la transacción	77
4.3.3.1.4.	Establecer los atributos que definen cada entidad, modelando sus puntos de inicio	78
4.3.3.1.5.	Establecer los atributos que hacen parte de la transacción a modelar como relaciones entre entidades	78
4.3.3.2.	Establecer instrumentos de documentación de elementos para el modelamiento	79
4.3.3.3.	Diagramar y socializar modelos de datos	79
4.3.3.3.1.	Diagramar el modelo conceptual y lógico	79
4.3.3.3.2.	Socializar los diagramas con proveedores y usuarios de la dimensión	79
4.3.3.3.3.	Actualizar permanentemente los modelos	79
4.3.4.	Salidas	79
4.3.5.	Usuarios	80
5. INTEGRACIÓN E INTEROPERABILIDAD DE DATOS		81
5.1. Conceptualización.....		81
5.1.1.	Integración e interoperabilidad de datos: definición	81
5.1.2.	Integración e interoperabilidad de datos: elementos	85
5.1.3.	Integración e interoperabilidad de datos: estándar del lenguaje común de intercambio	86
5.2. Propósito		87
5.2.1.	Metas	87
5.2.2.	Alcance en la fase II	88
5.2.3.	Punto de referencia para la interoperabilidad	88
5.3. Desarrollo de la fase II de integración e interoperabilidad de datos.....		89
5.3.1.	Proveedores	90
5.3.2.	Entradas	91
5.3.3.	Actividades	91
5.3.3.1.	Asignar responsables del intercambio de la información de Interoperabilidad	91
5.3.3.2.	Identificar los procesos susceptibles a Interoperar	91
5.3.3.3.	Definir los requerimientos de integración e interoperabilidad y el ciclo de vida de los datos	92

5.3.3.4. Identificar oportunidades de mejora en los procesos implementando servicios de intercambio de información	93
5.3.3.5. Definir servicios de intercambio de la información proveedor/consumidor	93
5.3.3.6. Divulgar los servicios de intercambio de la información al interior de la entidad	93
5.3.3.7. Publicar servicios de intercambio de la información en el Directorio de Servicios	93
5.3.4. Participantes	93
5.3.5. Salidas	94
5.3.6. Usuarios	95
6. GESTIÓN DE DATOS MAESTROS Y DE REFERENCIA	96
6.1. Conceptualización	96
6.1.1. Datos maestros	96
6.1.2. Datos de referencia	96
6.1.3. Estructura de datos de referencia	96
6.1.3.1. Lista simple	96
6.1.3.2. Listas de referencia cruzada	97
6.1.3.3. Taxonomías	97
6.1.4. Estándares del dominio de los datos	98
6.2. Propósito	98
6.3. Metas	99
6.4. Alcance fase II	99
6.5. Desarrollo de la fase II de gestión de datos maestros y de referencia	99
6.3.1. Proveedores	100
6.3.2. Entradas	100
6.3.3. Actividades	100
6.3.3.1. Implementar la metodología de identificación de datos maestros y de referencia	101
6.3.3.1.1. Identificar datos maestros y de referencia	101
6.3.3.1.2. Establecer datos maestros y catalogar datos de referencia	101
6.3.3.2. Establecer instrumentos de documentación de datos maestros y de referencia	102
6.3.3.2.1. Datos maestros	102
6.3.3.2.2. Datos de referencia	102
6.3.3.3. Incluir los datos maestros y de referencia en el modelo de datos y el repositorio de metadatos	103
6.3.4. Salidas	104
6.3.5. Usuarios	104
7. GESTIÓN DE METADATOS	106
7.1. Conceptualización	106
7.1.1. Tipologías de metadatos	107
7.1.2. Categorías de metadatos	111
7.2. Propósito	112
7.2.1. Metas	113
7.2.2. Alcance fase II	113
7.3. Desarrollo de la fase II de gestión de metadatos	113
7.3.1. Proveedores	115

7.3.2. Entradas	115
7.3.3. Actividades	115
7.3.3.1. Entender los requerimientos de metadatos	116
7.3.3.1.1. Principios para requerimientos de metadatos	116
7.3.3.1.2. Metas de los requerimientos	120
7.3.3.2. Definir la estrategia de metadatos	123
7.3.3.2.1. Política de gestión de metadatos	123
7.3.3.2.2. Identificar fuentes de metadatos	124
7.3.3.2.3. Entrevistas	124
7.3.3.2.4. Establecer una base de conocimiento de negocio y técnica	124
7.3.3.2.5. Búsqueda de estándares	124
7.3.3.2.6. Documentación	125
7.3.3.2.7. Medir el uso y la eficacia de la estrategia	126
7.3.3.3. Proponer instrumentos de documentación de metadatos	128
7.3.4. Salidas	130
7.3.5. Usuarios	131
8. CALIDAD DE DATOS	132
8.1. Conceptualización.....	132
8.1.1. Definición de la calidad de datos	132
8.1.2. Características y metodología para la construcción de indicadores de calidad	134
8.1.3. Tipología de indicadores según la Guía de Cuantificación, Medición y Seguimiento a la Gestión de Procesos (CMSP) de la Universidad Nacional de Colombia	140
8.1.4. Criterios o características de los indicadores según la Guía de CMSP de la Universidad Nacional de Colombia	141
8.1.5. Pasos metodológicos para la construcción de indicadores	142
8.2. Propósito	146
8.2.1. Metas	146
8.2.2. Alcance fase II	147
8.3. Desarrollo de la fase II de calidad de datos	147
8.3.1. Proveedores	148
8.3.2. Entradas	148
8.3.3. Actividades	148
8.3.3.1. Propuesta de indicadores de calidad para el momento inicial del ciclo de vida del dato	148
8.3.3.2. Socialización y ajuste de indicadores de calidad para el momento inicial del ciclo de vida del dato	154
8.3.3.3. Priorizar datos para realizar evaluación de calidad	155
8.3.4. Salidas	155
8.3.4.1. Formulación de indicadores que midan la calidad de los datos en la etapa inicial del su ciclo de vida, o captura de la información, junto con su instructivo de construcción	155
8.3.4.2. Ficha técnica de los indicadores de calidad de los datos en la etapa inicial de su ciclo de vida	157
8.3.4.3. Procedimiento para la implementación de los indicadores de calidad propuestos	157

9. CONCLUSIONES Y RECOMENDACIONES.....	158
Referencias	159
Glosario	163

LISTA DE FIGURAS

Figura 1 Dimensiones de la gestión de datos en la metodología DAMA.....	15
Figura 2 Proceso para la gestión del riesgo de la UNAL.....	35
Figura 3 Marcos operativos de gobierno de datos.....	39
Figura 4 Modelo de gobierno de los datos.....	40
Figura 5 Diagrama de Carriles (Swimlane).....	54
Figura 6 Componentes del diagrama de flujo de datos, notación Yourdon-De Marco	54
Figura 7 Mapa de Procesos de la Universidad Nacional de Colombia.....	56
Figura 8 Diagrama de flujo de datos nivel 0.....	64
Figura 9 Sección de la Matriz CRUD.....	66
Figura 10 Tipos de modelos de datos y ejemplo, adaptación de Ansari, (2017)	69
Figura 11 Representación gráfica de entidades en el modelo conceptual	70
Figura 12 Representación de relaciones en el modelo conceptual	70
Figura 13 Representación de cardinalidad y ordinalidad del modelo conceptual	70
Figura 14 Representación de herencia en el modelo conceptual	71
Figura 15 Representación de agregación o composición en el modelo conceptual	71
Figura 16 Representación de entidades y relaciones en el modelo lógico	72
Figura 17 Representación de entidades y relaciones en el modelo físico	73
Figura 18 Arquitectura tecnológica típica de un modelo de inteligencia de negocios – integración.	
Fuente: Rodríguez y Bernal (2019)	82
Figura 19 Esquema de interoperabilidad de datos. Fuente: https://dntic.unal.edu.co/index.php/proyectos/interoperabilidad	83
Figura 20 Modelo conceptual del marco de interoperabilidad MinTIC. Fuente: https://www.mintic.gov.co/arquitecturati/630/w3-propertyvalue-8117.html	84
Figura 21 Diagrama de arquitectura de interoperabilidad de datos del estándar de integración de datos e interoperabilidad. Fuente: http://lenguaje.mintic.gov.co/aprende-del-lenguaje	86
Figura 22 Interoperabilidad dentro del diseño de arquitectura de datos	89
Figura 23 Tipologías por capa para metadatos.....	107
Figura 24 Ejemplo de identificación de metadatos de valor y estructura para la codificación de departamentos de Colombia DIVIPOLA.....	108
Figura 25 Ejemplo de identificación de metadatos de formato y estructura para la codificación de departamentos de Colombia DIVIPOLA.....	109
Figura 26 Ejemplo de identificación de metadatos de intercambio para la codificación de departamentos de Colombia DIVIPOLA.....	110
Figura 27 Vista detallada de cada una de las capas de metadatos.....	110
Figura 28 Representación de la categorización de metadatos	112
Figura 29 Mapa conceptual de la gestión de la ruta en la gestión de metadatos	114
Figura 30 Ficha técnica sobre el índice de Resiliencia Epidemiológica Municipal - parte 1 (Ministerio de Salud, 2021)	118
Figura 31 Ficha técnica sobre el índice de Resiliencia Epidemiológica Municipal - parte 1 (Minsalud, 2021)	119
Figura 32 Representación de metas para requerimientos de metadatos	121
Figura 33 Analogía de estrategia de metadatos vs. Cubo Rubik.....	126
Figura 34 Niveles de madurez para la estrategia de metadatos	127
Figura 35 Gobierno de datos, administración y servicios. Guía de fundamentos para la gestión de datos (DAMA, 2010, p. 43)	134
Figura 36 Causas de problemas de calidad de datos – perspectiva organizacional corporativa. Fuente: Rodríguez (2020, p. 7).	135

Figura 37 Pasos metodológicos para la construcción de indicadores. Fuente: Universidad Nacional de Colombia (2012, p.19) 142
Figura 38 Clasificación de problemas de calidad de datos. Fuente: Rodríguez (2020, p. 6) 152

LISTA DE TABLAS

Tabla 1 Principios generales de protección de datos. Tomado de: DAMA, 2017, p. 56	21
Tabla 2 SIPOC para el desarrollo de la dimensión de ética de datos	24
Tabla 3 Niveles de madurez para la gestión ética de los datos	28
Tabla 4 Niveles de madurez para la gestión ética de los datos desagregados por componentes	28
Tabla 5 Órganos de gobierno. Adaptado de: DAMA (2017, p. 76)	43
Tabla 6 SIPOC para el desarrollo de la dimensión de gobierno de datos	46
Tabla 7 Notación básica utilizada en el Diagrama Entidad-Relación	53
Tabla 8 SIPOC para arquitectura de datos aplicado a la Universidad	56
Tabla 9 Macroprocesos y procesos identificados para el análisis de necesidades - Fase I	58
Tabla 10 Paralelo entre la metodología TOGAF-ADM y la metodología DAMA y su articulación con la arquitectura empresarial	60
Tabla 11 Ejemplo de necesidades de la información por categorías a nivel de proceso (Fase 1 del proyecto)	65
Tabla 12 Símbolos con su significado, utilizados en el modelo lógico	72
Tabla 13 Ejemplos de tipos de datos presentes en el modelo físico	73
Tabla 14 Macroprocesos, procesos y sistemas de la información enmarcados en el alcance de la fase I para la dimensión de modelado y diseño de datos	74
Tabla 15 SIPOC de la dimensión de modelado y diseño de datos	75
Tabla 16 SIPOC para Integración e Interoperabilidad de datos en la UNAL	90
Tabla 17 Fragmento de lista de referencia simple. Ejemplo de códigos de bloqueo (SIA)	96
Tabla 18 Lista de referencia cruzada. Ejemplo de equivalencia entre códigos de sede para SIA e ISYS	97
Tabla 19 Taxonomía. Ejemplo fragmento de líneas de investigación sistema de la información de admisiones ISYS	98
Tabla 20 SIPOC de la dimensión de gestión de datos maestros y de referencia fase II	99
Tabla 21 Ítems para identificar datos maestros y de referencia	101
Tabla 22 Ítems para clasificar datos de referencia	102
Tabla 23 Preguntas de contextualización para metadatos (Burbank, 2020)	111
Tabla 24 SIPOC para la dimensión de gestión de metadatos en la fase II	113
Tabla 25 Comparativo de fuentes	125
Tabla 26 Definiciones términos clave ISO 11179 parte 3	129
Tabla 27 Atributos mínimos para elementos de metadatos según la norma ISO 11179 parte 3	130
Tabla 28 Dimensiones comunes de la calidad de datos. Adaptado de: DAMA (2017, pp. 435-437)	136
Tabla 29 Dimensiones de calidad según la norma ISO/IEC 25012. Fuente: Norma ISO/IEC 25012 (2019) 138	
Tabla 30 Criterios a valorar en la selección de indicadores. Universidad Nacional de Colombia (2012, p.22)	144
Tabla 31 SIPOC para el desarrollo de la fase II de calidad de datos	147
Tabla 32 Características del diccionario de datos de cada dependencia. Fuente: construcción propia 149	
Tabla 33 Métricas básicas de propósito general para datos cuantitativos	150
Tabla 34 Métricas para dimensiones inherentes asociadas a datos cualitativos	150
Tabla 35 Relación entre dimensiones y problemas de calidad – Objeto de medición: DATOS. Fuente: Rodríguez (2020, p. 65)	152
Tabla 36 Listado de indicadores para aplicar transversalmente a las dependencias de la UNAL	155

LISTA DE ANEXOS

- Anexo 1** Cruce de la información disponible y necesidades existentes
- Anexo 2** Diagrama de entidades maestras
- Anexo 3** Matriz Necesidades de Información
- Anexo 4** Matriz CRUD Identidades – Procesos
- Anexo 5** Formulario para el levantamiento de datos e información
- Anexo 5B** Instructivo de diligenciamiento para el formulario de levantamiento de datos e información
- Anexo 5C** Propósito de las secciones del formulario para el levantamiento de datos e información
- Anexo 6** Prototipo de modelo conceptual
- Anexo 7** Prototipo de modelo lógico
- Anexo 8** Instrumento de documentación de datos e información
- Anexo 9** Formato para lista de referencia simple
- Anexo 9A** Instructivo de diligenciamiento del formulario para la documentación de tablas de referencia simple
- Anexo 9B** Propósito de las secciones del formulario para la documentación de tablas de referencia simple
- Anexo 10** Formato para lista de referencia cruzada
- Anexo 10A** Instructivo de diligenciamiento del formulario para la documentación de tablas de referencia simple cruzada
- Anexo 10B** Propósito de las secciones del formulario para la documentación de datos de referencia cruzada
- Anexo 11** Formato para documentación de metadatos
- Anexo 11A** Instructivo de diligenciamiento de formato de documentación de metadatos
- Anexo 11B** Propósito de formato de documentación de metadatos

INTRODUCCIÓN

Actualmente La Universidad Nacional de Colombia no cuenta con una política institucional que guíe la gestión del dato y la información requerida a nivel institucional. Por ello, mediante el Oficio R-566 de 2019 se constituyó la Rectoría de CLÚSTER y GRUPOS TIC de la Universidad Nacional de Colombia para la Transformación Digital, dentro de ellos, el Grupo de Gestión de la Información de la Universidad Nacional de Colombia. Este grupo es el encargado de liderar el Proyecto BPUN 499, Gestión de la Información en la Universidad Nacional de Colombia Fase I, con el objetivo de definir un modelo y formular la política para la gestión del dato y la información en la Universidad Nacional de Colombia con visión al año 2034.

Como parte de las actividades de este proyecto, el Grupo de Gestión de la Información desarrolló un diagnóstico de la Política de Gestión de la Información (2021) a través del cual se identificaron problemáticas existentes en la Universidad relacionadas con el manejo de la información. Principalmente se encontró una débil capacidad para usar, apropiar y aprovechar la información interna y externa de la Universidad para el mejoramiento de la toma de decisiones, la transparencia institucional y la rendición de cuentas. Algunas de las causas más recurrentes que generan esta situación general son:

- Debilidad en la calidad de la información.
- El desarrollo de prácticas de seguimiento y evaluación que no son útiles para el mejoramiento en la toma de decisiones.
- La debilidad en el gobierno estratégico y la administración de la información.
- La débil orientación de la gestión y los sistemas de la información hacia la satisfacción de las necesidades de los usuarios.
- El débil diseño institucional para la gestión de la información.
- La debilidad de los sistemas de la información en cuanto a infraestructura y oferta de servicios

Por todo lo anterior, la primera fase del presente proyecto busca, por un lado, consolidar una política institucional que guie la gestión y disposición de la información requerida a nivel institucional con la que se busca, entre otros:

- Definir los principios que rigen la función de la gestión de la información.
- Establecer articulaciones con políticas y normatividad interna relacionada con la gestión de la información.
- Establecer el Comité de gobierno de la información, así como los roles, responsables, custodios, administradores y gestores de datos.
- Unificar las normativas para la gestión de la información.

Por otro lado, en esta primera fase se establece un modelo probado que, partiendo de las necesidades identificadas mediante el diagnóstico de gestión de la información, orientará la gestión de la información a través de normas, actores, herramientas, instrumentos y procesos requeridos para la disposición de la

información institucional con altos estándares de calidad, de manera oportuna y con la gobernabilidad requerida.

Para la constitución del modelo se abordó la metodología del *Data Management Institute* (DAMA) recopilado en la publicación Guía del conocimiento para la gestión de datos (2017). Allí se presentan 11 áreas de conocimiento sobre los datos que en el presente proyecto se denominan *dimensiones*, las cuales son complementarias y dependientes entre sí (**Figura 1**). Para la Fase I del presente proyecto, el Modelo de Gestión de la Información aborda las siguientes dimensiones: arquitectura, calidad, ética, modelado y diseño de datos, gestión de metadatos, gestión de datos maestros y referencia, gobierno de datos e interoperabilidad.

Figura 1 Dimensiones de la gestión de datos en la metodología DAMA



IT4+ del Ministerio de Tecnologías de la Información y las Comunicaciones (2019) buscando afinamiento en detalles específicos de implementación.

Metodología DAMA

Cada dimensión del modelo desarrollado en los capítulos siguientes presenta un diagrama SIPOC. Este hace referencia al acrónimo en inglés *Supplier – Inputs- Process- Outputs – Customers*, que en su traducción al español significa Proveedores, Insumos, Proceso, Resultados y Clientes; “Los mapas SIPOC proporcionan un resumen amplio de los elementos clave en el proceso y explican quién es el propietario de este, cómo se adquieren los insumos, quién atiende el proceso y cómo se agrega valor” (Evans & Lindsay, 2015, p. 483). Esta es una herramienta incorporada en las metodologías desarrolladas en el ámbito de la gestión de calidad, con su apogeo en los años 80 en occidente, luego de su desarrollo en oriente. Esta herramienta es utilizada para la gestión y mejora de procesos ayudando a entender mejor el proceso mediante un visión global y sencilla del mismo.

A continuación, se hace una descripción de los elementos que componen esta herramienta:

- **Proveedor (*Supplier*):** persona, organización o proceso que aporta las entradas (insumos) al proceso. Los proveedores pueden ser internos o externos.
- **Entradas (*Input*):** todo lo que se requiere para el desarrollo de cada una de las actividades inmersas en el proceso. Se consideran recursos a la información, documentos, informes, materiales, talento humano, financieros, entre otros.
- **Proceso (*Process*):** conjunto de actividades y tareas que transforman los insumos (entradas) en salidas (productos o servicios), logrando un valor agregado.
- **Salida (*Output*):** productos o servicios resultado del proceso.
- **Cliente (*Customer*):** persona, organización o proceso que recibe los productos o servicios (salidas) del proceso. Así como los proveedores, los clientes o usuarios pueden ser internos o externos a la organización. Es pertinente precisar que, para organizaciones gubernamentales, el enfoque cambia un poco en relación con la concepción de clientes, ya que las entidades públicas buscan desarrollar sus actividades para los ciudadanos, como usuarios de sus procesos, buscando el disfrute pleno de sus derechos (Modelo Integrado de Planeación y Gestión - MIPG).

El SIPOC es una herramienta utilizada por varias de las metodologías de gestión de calidad tales como el *Lean Six Sigma*, DMAIC, *Kaizen* e incluso en la implementación de la ISO 9001. Junto con los diagramas de flujo, el VSM, diagrama de hilos o *Genchi Genbutsu* se constituyen en elementos que permiten a los analistas conocer a diferentes niveles de profundidad los procesos, identificando actividades o tareas susceptibles a evaluación y mejora.

Esta herramienta es también llevada al mundo de la gestión de los datos y la información en el marco de referencia utilizado por el DAMA en el DMBook, adaptándolo al diagrama de contexto para identificar a nivel de proceso los actores, insumos, actividades y productos que intervienen en la gestión en cada una de las áreas de conocimiento que conforman el alcance general de la gestión de datos (DAMA Internacional, 2017).

La Universidad ha definido utilizar el SIPOC como herramienta de apoyo para el desarrollo de las diferentes dimensiones con las cuales se abordará la gestión de los datos y la información en la institución, al reconocer valor agregado que genera, estandarizando la manera en la cual se abordará cada uno de los temas, brindando al lector una estructura bajo la cual le permitirá un mejor entendimiento de los temas tratados, estructura que se mantendrá a lo largo del desarrollo del modelo.

Contenidos del modelo

El presente documento busca establecer lineamientos procedimentales y metodológicos para una efectiva gestión de la información en la Universidad. Los fundamentos conceptuales y metodológicos aquí incluidos se inspiran en el marco de referencia DAMA International (2017) y en el Modelo IT4+ del MINTIC. Se incluyen las dimensiones de manejo ético de datos, gobierno de los datos y la información, arquitectura de datos, modelado y diseño de datos, integración e interoperabilidad de datos, gestión de datos maestros y de referencia, gestión de metadatos y calidad de los datos.

En cada una de estas dimensiones se desarrollan los elementos conceptuales básicos, el propósito y las metas de la dimensión y los lineamientos procedimentales y metodológicos a implementar en la segunda fase de los desarrollos del Modelo de Gestión de la Información (la primera fase correspondió a diseñar la Política de Gestión de la Información, el presente documento, y sus respectivos anexos). Las dimensiones seleccionadas y que componen la estructura del documento son las siguientes:

- Ética para el manejo de datos
- Gobierno de los datos y la información
- Arquitectura de datos
- Modelamiento y diseño de datos
- Integración e interoperabilidad de datos
- Gestión de datos maestros y de referencia
- Gestión de metadatos
- Calidad de los datos

Estas dimensiones corresponden a una priorización realizada por parte del Clúster de Gestión de la información de la Universidad entre los años 2020 y 2021. En este periodo de tiempo se llevaron a cabo una serie de mesas técnicas entre representantes de las áreas y sistemas de la información más relevantes dentro de la organización para poder realizar diseños institucionales pertinentes, esta estructura es producto de este trabajo conjunto.

Dimensiones no abordadas

Se debe tener claridad que la gestión de los datos y la información constituye un proceso de actualización permanente, puesto que constituye una rama del conocimiento en permanente evolución y que debe ser adaptada a la coyuntura organizacional de la Universidad. Las dimensiones aquí desarrolladas consisten un primer trabajo conjunto y priorización, sin embargo, el DAMA International

(2017) ha identificado otras dimensiones que también requieren un desarrollo e implementación de lineamientos procedimentales y metodológicos, estas dimensiones restantes son:

- Almacenamiento de datos y operaciones
- Seguridad de datos
- Gestión de documentos y contenido
- Data warehousing e inteligencia de negocio
- Big data y ciencia de datos
- Evaluación de la madurez de la gestión de datos
- Organización de la gestión de datos y expectativas de roles
- Gestión de datos y gestión del cambio organizacional

Para poder implementar estas dimensiones es necesario contar con avances significativos en las dimensiones seleccionadas en el presente documento. Sin estos desarrollos previos no se contarán con elementos esenciales en la Universidad para poder obtener los beneficios que la implementación de estas dimensiones podría brindar. En la fase II de la implementación de la Política y el Modelo de Gestión de la Información se debe avanzar en el perfeccionamiento de los presentes lineamientos procedimentales y metodológicos a través de su pilotaje y de la retroalimentación que se pueda obtener posterior a su implementación. Luego de dicha retroalimentación se podrá valorar el direccionamiento estratégico más pertinente.

1. ÉTICA

1.1. Conceptualización

Según lo establecido en la Guía del Conocimiento para la Gestión de Datos, la ética se define como “un conjunto de principios de conducta basados en ideas de lo correcto y lo incorrecto” (DAMA Internacional, 2017, p. 51), lo que induce a identificar como principios éticos a aspectos de justicia, responsabilidad, integridad, transparencia, igualdad, entre otros.

Los principios éticos deben estar presentes en todo el ciclo de vida del dato: desde su captura (almacenamiento y tratamiento) hasta su eliminación o archivo. Estos principios otorgan una buena reputación al encargado del tratamiento de datos, constituyen un mecanismo preventivo de uso indebido de la información y apoyan la implementación de los requerimientos legislativos vigentes relacionados con *habeas data* y la Ley de Transparencia¹.

Actualmente la Universidad Nacional de Colombia tiene un Compromiso Ético, adoptado mediante la Resolución 11 de 2018, que induce a todos los integrantes de la comunidad universitaria (estudiantes, servidores públicos, docentes y administrativos) a cumplir con el ejercicio de sus funciones, basándose en la aplicación de principios éticos, integrales y transparentes. A partir de los principios de pertenencia, equidad, honestidad, solidaridad, responsabilidad, respeto y diálogo, en esta dimensión, se busca definir como algunos de estos principios deben ser entendidos dentro de las actividades relacionadas con el manejo de información.

Lo anterior, para efectos de ser concordantes con la tarea de apropiación y sensibilización de los principios éticos aplicados al manejo de la información debe realizarse en coordinación con la Dirección Nacional de Personal Académico y Administrativo, la cual está a cargo de “promover la implementación de programas, estrategias y actividades que contribuyan a la divulgación y socialización del Compromiso Ético” (2018, Art. 3, numeral 3). La tarea de apropiación y sensibilización de los principios éticos aplicados al ciclo de vida del dato incluye:

- Capacitación permanente a los funcionarios.
- Identificación, prevención y mitigación de acciones en torno a los riesgos de seguridad, y tratamiento de la información conforme a la normatividad vigente.
- Suscripción de unas cláusulas éticas específicas para los profesionales que gestionan datos y cuyo compromiso es propender por el bien de la sociedad, más allá de su objeto contractual.

¹ Ley 1581 de 2012 y Ley 1712 de 2014.

Vale la pena aclarar que la aplicación de los principios éticos está atada a la existencia de controles que validen su cumplimiento. Para ello, en la siguiente sección se presentan los avances de la Universidad en materia legal para protección de datos.

1.1.1. Relación entre los principios éticos de la Universidad Nacional y el manejo de datos

De acuerdo con la Resolución 11 de 2018 y su anexo, las actividades académicas, de investigación, de extensión y de tipo administrativo, deben ejecutarse bajo los siguientes principios, que a su vez se encuentran alineados con los principios administrativos de la Ley 1437 de 2011.

- **Diálogo:** medio para la solución de conflictos y herramienta para el intercambio de ideas.
- **Equidad:** valor de la igualdad en la dignidad y la justicia social, puesto en práctica a través de la aplicación de acuerdos éticos.
- **Honestidad:** manifestación de comportamientos sinceros, transparentes y coherentes; se manifiestan en “la capacidad de negarse a engañar o apropiarse de lo ajeno y de aceptar los propios límites para no mentirse a uno mismo y a los demás” (anexo 1 de la Resolución 11/18). Es un principio que tiene la capacidad de contrarrestar la presencia de la corrupción.
- **Pertenencia:** sentido de apropiación por una organización social enmarcado en la consecución de “metas, aspiraciones, valores e ideales” y acompañado con el concepto de inclusión, entendido como la expresión de una historia que se construye colectivamente.
- **Respeto:** comportamientos activos de afinidad entre individuos que incluyen acciones de tolerancia, comprensión, convivencia, buen trato y aceptación.
- **Responsabilidad:** capacidad y voluntad de responder por las acciones y consecuencias propias de cada individuo bajo un entorno de salud y prevención de acciones que puedan generar daño a un tercero.
- **Solidaridad:** capacidad de generar apoyo, cooperación y unión para la consecución de logros bajo intereses comunes.

La Resolución 11 de 2018 de la Universidad Nacional de Colombia es un referente de cómo hacer las cosas siguiendo un compromiso ético que orienta el fortalecimiento del carácter institucional y el cumplimiento de la misión de la Universidad. Además del compromiso ético, la Universidad debe dar cumplimiento a la normativa vigente en torno a la gestión ética de los datos, así el accionar de la Universidad se debe fundamentar en la Ley 1581 de 2012 la cual desarrolla jurídicamente los principios o formas de tratar los datos personales y en la Ley 1712 de 2014 en donde se define lo que se entiende por información pública o sensible, quiénes son los sujetos que pueden acceder a ella y en qué condiciones pueden acceder. La información que no se encuentre restringida en estas dos leyes o sus normas concordantes puede circular libremente entre cualquier ciudadano.

Estas dos normas son la guía para clasificar y tratar toda la información manejada y administrada por la Universidad, de modo que lo que no entre en esta clasificación no tiene restricciones jurídicas en su

tratamiento y, en ese sentido, todo ciudadano tiene derecho a acceder a la información sin restricción alguna, en cumplimiento a la legislación vigente en materia de acceso a la información y protección de datos.

Por otra parte, en la Guía del Conocimiento para la Gestión de Datos (DAMA Internacional, 2017) se describen los principios para el procesamiento de datos (**Tabla 1**) que corresponden con los estipulados por la Organización para la Cooperación y el Desarrollo Económicos (OCDE). Estos principios se encuentran contenidos en el reglamento general de protección de datos del Consejo y Parlamento Europeo (2016). No obstante, en la guía del DAMA se hace énfasis en la aplicación de la normatividad vigente de cada país para dar protección y seguridad al manejo de los datos.

Tabla 1 Principios generales de protección de datos. Tomado de: DAMA, 2017, p. 56

Principio	Descripción del principio
Justicia, lealtad, transparencia	Los datos personales deberán ser procesados de forma lícita, justa y en transparencia con la persona afectada.
Limitación del propósito	Los datos personales deben ser recolectados para fines específicos, explícitos y legítimos y no procesados de una manera que sea compatible con esos propósitos.
Minimización de datos	Los datos personales deben ser adecuados, pertinentes y limitados a lo que sea necesario con los fines para los cuales se procesan.
Exactitud	Los datos personales deben ser exactos y cuando sea necesario, mantenerse actualizados. Se deben tomar todas las medidas razonables para garantizar que los datos personales inexactos (dependiendo del propósito para el que se procesan) sean borrados o rectificados sin demora.
Limitación de almacenamiento	Los datos deberán conservarse en un formato que permita identificar a los sujetos (individuos) durante un período no mayor al necesario para los fines para los que se procesaron los datos personales.
Integridad y confidencialidad	Los datos deben ser tratados de manera que garantice la seguridad adecuada de los datos personales, incluida la protección contra el procesamiento no autorizado o ilícito y contra la pérdida accidental, destrucción o daño, utilizando medidas técnicas u organizaciones apropiadas.

1.1.2. Situaciones no controlables ante la ausencia de principios éticos en el manejo de datos

Son cinco las situaciones que pueden ocurrir ante la ausencia de un lineamiento ético en el manejo de los datos: 1) presentación de la información a conveniencia, 2) representación engañosa de los datos, 3) presentación de definiciones conceptuales confusas o incompletas, 4) realización de comparaciones

no válidas, 5) presentación de estimaciones de datos sesgadas. A continuación, se presentan las características de cada una de ellas.

1.1.2.1. Presentación de la información a conveniencia

Cuando de forma intencional se excluye información verídica, se incluye información falsa o se cambia parcialmente información en algunas variables. Esto puede ocurrir por:

- Obtención de algún beneficio para la Universidad o para alguna dependencia en particular.
- Descuido en la adecuada captura, tratamiento, almacenamiento, uso, circulación o eliminación de los datos.
- Carencia de una fuente oficial que reporte los datos.
- Migración de la información sin la aplicación de protocolos e indicadores que midan y den cuenta del éxito del proceso.
- Falta de políticas de calidad y de uniformidad de procedimientos en el tratamiento de la información.
- Falta de memoria institucional sobre la aplicación y los resultados de auditorías aplicables a la calidad de los datos.
- Falta de documentación histórica que soporte los tratamientos y transformaciones aplicados a los datos, así como la posibilidad de recuperar la información original.
- Falta de herramientas de gestión de riesgo articulados con los sistemas, controles, instrumentos y metodologías, entre otros, que faciliten los procesos de prevención y mitigación de impactos negativos o adversos detectados durante cada una de las etapas del ciclo de vida del dato.
- Falta de campañas de sensibilización para conservar la información bajo las condiciones de seguridad necesarias para evitar su pérdida, uso o acceso no autorizado o fraudulento, de manera que se garantice la disponibilidad, integridad y confidencialidad de los datos personales.
- Carencia de políticas en medidas de seguridad que se apliquen para minimizar los riesgos asociados al intercambio de datos personales bien sea de manera física o electrónica, como interceptación, consulta no autorizada, fraude, pérdida o robo de la información.

1.1.2.2. Representación engañosa de los datos

- Se comparan dos variables en diferentes unidades. Por ejemplo, la comparación entre la desviación estándar de las notas de un curso A, con la varianza de un curso B, parámetros estadísticos no comparables. También sucede cuando se comparan las desviaciones estándar de dos cursos: uno con escala de calificación de 1 a 5 y el otro de 1 a 100.
- Se comparan modelos o situaciones que, aunque tienen propósitos de medición final iguales, internamente se obtienen bajo metodologías diferentes o se aplican a poblaciones no comparables. Por ejemplo, comparar resultados de Pruebas Saber con resultados de Pruebas Pisa, las cuales tienen metodologías diferentes y se aplican a poblaciones de condiciones demográficas diferentes.

1.1.2.3. Presentación de estimaciones de datos segadas

Esto sucede al estimar el valor o parámetro estadístico con una muestra que lleva a un resultado que no representa toda la población; por ejemplo, indicar que el promedio de todos los estudiantes de la facultad de ciencias es 4.5, excluyendo intencionalmente a tres carreras, que hacen que ese promedio eventualmente suba a 4.7 o disminuya a 4.0, por lo que 4.5 sería una estimación sesgada.

1.1.3. Metas

- Definir los principios éticos de la Universidad Nacional en torno al manejo de datos.
- Definir un diagnóstico de cumplimiento normativo con respecto a la seguridad de la información y el manejo de datos.
- Construir un documento que describa los mecanismos de sensibilización presentes en la Universidad frente al manejo de los datos junto con sus estrategias de verificación de cumplimiento.

1.2. Propósito

Teniendo en consideración que la conceptualización de ética de los datos induce a identificar como principios éticos a aspectos de justicia, responsabilidad, integridad, transparencia, igualdad, entre otros. En esta dimensión se fijarán los principios éticos bajo los cuales debe dársele tratamiento a la información y deben estar presentes en todas las etapas del ciclo de vida del dato, ya que son reglas del actuar de la comunidad de la Universidad Nacional frente al manejo del dato.

1.2.1. Alcance Fase I

El alcance de la fase I (correspondiente al diseño e implementación del Proyecto 499 – *Gestión de la Información en la Universidad Nacional de Colombia*) se orientó a la definición de los principios éticos de la Universidad Nacional de Colombia enmarcados en el ciclo de vida del dato. Estos serán insumo para fijar los lineamientos y mecanismos de sensibilización para el proceso de apropiación en la Universidad. Así las cosas, el desarrollo de esta primera fase, se direccionó hacia la revisión de los principios establecidos en el compromiso ético y la realización del diagnóstico de mecanismos de sensibilización presentes en la Universidad para socializar los principios éticos aplicables al manejo de datos.

Vale la pena aclarar que los principios éticos sobre el tratamiento de datos son de aplicación transversal a toda la información de la Universidad y no distinguen entre la clasificación que se le haga a un conjunto de datos, bien sea bajo un principio jurídico, una regla de investigación, un nivel jerárquico de la Universidad o cualquier otro esquema que agrupe o estratifique su información.

En fases posteriores se trabajará en las siguientes metas:

- Generar un diagnóstico sobre el estado actual del manejo ético de los datos y definición de una hoja de ruta o acciones de mejora para cerrar las brechas.
- Elaborar un mapa de los riesgos éticos en el manejo de los datos e identificación de acciones de mitigación.
- Fortalecer la cultura del manejo apropiado del dato acorde a los principios y valores éticos.
- Dar continuidad al manual de mecanismos de sensibilización e información frente al manejo de los datos junto con sus estrategias de verificación de cumplimiento.
- Plantear indicadores para monitorear y hacer seguimiento a la aplicación de acciones éticas en el manejo de datos.

Finalmente, en fases posteriores se considerarán las siguientes actividades:

- Plan de acción para el cierre de brechas en el manejo ético de los datos.
- Elaboración del mapa de riesgos éticos.

1.3. Desarrollo de la Fase II de ética

Tabla 2 SIPOC para el desarrollo de la dimensión de ética de datos

Proveedores	Entradas	Actividades	Salidas	Usuarios
Líderes de los procesos y de los sistemas de información DNED Vicerrectoría General	Política de Tratamiento de Datos Personales. Política de tratamiento y seguridad de los datos Compromiso ético de la UNAL	Realizar un diagnóstico de cumplimiento normativo para evaluar el estado de las prácticas institucionales actuales de la UNAL frente al manejo ético de los datos.	Documento de diagnóstico de cumplimiento normativo	Líderes de los sistemas de información Comité de Gobierno de la Gestión de Información DNED
Clúster de gestión de la información	Principios éticos definidos en la Fase I del proyecto	Socialización de los principios éticos para el manejo de datos y de información	Plan de socialización de los principios éticos.	Líderes de los sistemas de información Actores encargados de la gestión de datos y la información
Sistema Integrado de Gestión Académica, Administrativa y Ambiental	Marco Integral de la Gestión de Riesgos MIGR	Formulación de metodología para la identificación de los factores de riesgo frente al manejo ético de los datos	Metodología para la identificación de los factores de riesgo	Líderes de los sistemas de información DNED

1.3.1. Proveedores

Líderes de procesos y de los sistemas de información: personas que lideran los sistemas de la información y, por tanto, son quienes promueven la aplicación de prácticas adecuadas en la ejecución de sus labores cotidianas.

DNED: instancia creada mediante el Acuerdo 316 del Consejo Superior Universitario como un ente de gobierno encargado de asesorar, dirigir la formulación, coordinar y proponer lineamientos en materia tecnológica y de cultura digital.

Vicerrectoría General: es la instancia encargada de asesorar a los cuerpos colegiados en la formulación de propuestas para la definición de políticas en talento humano, desarrollo institucional, informática y comunicaciones, de recursos y servicios bibliográficos, y en general del sistema de gestión de la institución.

Clúster de gestión de la información UNAL: lugar donde se deben consignar los lineamientos asociados a buenas prácticas en el manejo de la información; en este caso, para indagar la existencia o no de principios éticos en el manejo de datos al interior de la Universidad.

Comité de gobierno de la información: Instancia principal de articulación interinstitucional encargada de los diseños programáticos y metodológicos en temas de gestión de la información de la Universidad Nacional de Colombia.

1.3.2. Entradas

Política de Tratamiento de Datos Personales: se establece mediante la Resolución 207 de 2021 en armonización con lo establecido en el artículo 15 de la Constitución Política de Colombia. Así pues, la política es entendida como lineamientos para establecer el debido proceso para la recolección, uso y tratamiento de datos personales.

Política para la Gestión de Información: política pendiente de aprobación, su objeto es constituirse como la principal directriz de la Universidad en materia de gestión de datos e información, orientada a desarrollar lineamientos institucionales para incrementar las capacidades de la UNAL en el uso, apropiación y aprovechamiento de la información.

Marco Normativo para la gestión de información: en la Fase I del presente proyecto se avanzó en la delimitación del marco normativo legal vigente en materia de gestión de información.

Compromiso ético de la UNAL: mediante la Resolución 11 de 2018 se adopta el Compromiso Ético de la Universidad Nacional de Colombia el cual es un instrumento que incluye un acuerdo en torno a los valores que deben guiar el comportamiento de la comunidad universitaria de la UNAL.

Principios éticos definidos como insumo del Modelo de Gestión de la Información: en la Fase I del presente proyecto se avanzó en la delimitación de los principios éticos que rigen la gestión de información.

Guía para la administración de riesgos de la UNAL: desde el Sistema de Gestión Académica, Administrativa y Ambiental Organizacional se diseñó la guía de administración de riesgos como un marco de referencia para gestionar riesgos siguiendo lo establecido en la norma ISO 31000:2009. En

ese sentido, el documento presenta las pautas metodológicas para administrar riesgos conforme a los lineamientos institucionales y en cumplimiento de la misión institucional de la Universidad.

1.3.3. Actividades

1.3.3.1. Realizar un diagnóstico de cumplimiento normativo para evaluar el estado de las prácticas actuales de la UNAL en torno al manejo ético de los datos

Las dependencias responsables de la gestión de datos e información deben tener suficiente conocimiento de las prácticas existentes en torno al manejo ético, así como las implicaciones que tiene para la Universidad no gestionar los datos y la información siguiendo lineamientos institucionales y principios éticos.

La actividad se estructura a partir de una serie de lineamientos procedimentales sustentados en el marco de referencia CMM (Capability Maturity Model), el cual describe cómo evolucionan las características de un proceso en una organización. Así pues, esta herramienta es útil para hacer una revisión estratégica del marco de gestión de datos, teniendo como referencia las características asociadas a los niveles de madurez.

A través de la ejecución de los pasos propuestos en esta sección, se espera que las dependencias tengan claridad de las acciones que deben ejecutar para evaluar el grado en que las prácticas actuales se alinean con los lineamientos normativos éticos.

Paso 1. Definir la normatividad que se va a evaluar según las necesidades de la dependencia.

Teniendo en cuenta que el marco normativo aplicable a la gestión ética de los datos en la Universidad es amplio y diverso, es preciso que la dependencia seleccione la normativa específica que desea evaluar tomando en consideración el cumplimiento de sus funciones en la Universidad.

De acuerdo con lo dispuesto en la técnica de análisis de brechas, es posible identificar las diferencias que existen entre las prácticas actuales de una organización y “lo que debería ser” es decir el escenario ideal (Ji y Kim, 2018). Por lo que la normatividad que seleccione la dependencia será tomada como el escenario ideal o el deber ser.

Por su parte, la elección de las prácticas a evaluar dependerá de las necesidades concretas de la dependencia. Estas pueden estar relacionadas con la transparencia en la gestión de información, la privacidad y seguridad de los datos, la responsabilidad en el uso de la información y demás prácticas asociadas con el manejo ético.

Algunos documentos que contienen directrices sobre el tratamiento ético de la información son:

- Manual Interno y procedimientos para la aplicación de la Política de Tratamiento de Datos Personales de la Universidad Nacional de Colombia.
- Compromiso ético de la Universidad Nacional de Colombia.
- Política de Tratamiento de Datos Personales de la Universidad Nacional de Colombia (Resolución 207 de 2021, Rectoría).
- Políticas sobre tratamiento y seguridad de datos, tanto a nivel nacional como a nivel interno de la Universidad.

Debido a que el interés es construir una herramienta de recolección de la información tipo test, es necesario que, de la normativa seleccionada, la dependencia elija un conjunto de artículos, controles o lineamientos concretos, con el fin de delimitar el marco de recolección de información. Estos artículos, controles o lineamientos concretos deben funcionar, en términos evaluativos y de recolección de datos, como las variables que van a ser medidas, analizadas e interpretadas en las dependencias de la UNAL.

Paso 2. Construir la herramienta de recolección de información.

La dependencia deberá construir una serie de preguntas para indagar sobre el cumplimiento en la aplicación de controles o en implementación de lineamientos propuestos en la normativa seleccionada. Estos son algunos ejemplos de preguntas que pueden orientar en esta tarea.

- ¿Se han asignado los roles y responsabilidades para la implementación de la Política de Tratamiento de Datos Personales? (**Gobernanza institucional**)
- ¿Se ha definido y aplicado un proceso de gestión de riesgos asociados a la transparencia en el uso de la información? (**Proceso de gestión de riesgos**)
- ¿Los planes de capacitación del personal abordan temáticas de cultura ética de en el uso, recolección e intercambio de información? (**Habilidades y conocimientos**)
- ¿La organización implementa herramientas de seguimiento y evaluación para identificar prácticas inadecuadas en el tratamiento ético de la información? (**Situación legal y cumplimiento**)

Para darle un orden al cuestionario, se propone agrupar los artículos según la correspondencia que tengan con las seis temáticas propuestas en el modelo de madurez de los datos propuesto por The ODI.

- **Gobernanza institucional y supervisión interna:** se refiere a la necesidad de que las organizaciones tengan definida una estrategia clara en torno al manejo ético de los datos, así como responsabilidades definidas para la ejecución de la estrategia.
- **Habilidades y conocimientos:** la organización gestiona actividades en torno a la promoción una cultura ética sobre el manejo de los datos, tales como capacitaciones e intercambio de conocimientos.
- **Procesos de gestión de riesgos:** La organización identifica, evalúa y establece controles frente a los factores de riesgos en torno al manejo ético de los datos.
- **Financiamiento y adquisiciones:** la organización invierte en la incorporación de prácticas éticas de datos.

- **Compromiso de las partes interesadas y del personal:** la organización se involucra con las partes involucradas en la recolección, uso e intercambio de datos y de información.
- **Situación legal y cumplimiento:** la organización cumple la normativa para evitar impactos negativos en la recolección, uso, intercambio de datos.

Paso 3. Aplicar el cuestionario con los actores estratégicos

Para la aplicación de la técnica de brechas es aconsejable utilizar un modelo de madurez como medio para evaluar el grado de implementación de la normativa en las prácticas de la Universidad Nacional. Los modelos de madurez proponen una serie niveles de cumplimiento.

En la aplicación del cuestionario se debe tener presente que los niveles propuestos no son el objetivo, sino el medio para evaluar el grado de implementación. El cuestionario se aplicará en mesas de trabajo con los actores relevantes en el uso e intercambio de la información y datos.

Tabla 3 Niveles de madurez para la gestión ética de los datos

Nivel de madurez	Peso porcentual	Descripción del nivel de madurez
INICIAL (I)	0%	Los procesos deseables son inexistentes, sin supervisión organizacional.
REPETIBLE (R)	25%	Los procesos se están volviendo refinados y repetibles, pero solo con un alcance de equipos o proyectos individuales, sin un alcance organizacional.
DEFINIDO (D)	50%	Procesos estandarizados en la organización en función de mejorar las prácticas.
GESTIONADO (G)	75%	Aplicación de procesos de monitoreo y aplicación de métricas de evaluación.
OPTIMIZADO (O)	100%	La organización intenta refinar y optimizar su proceso para hacer sus prácticas más eficientes.

Paso 4. Interpretación de los resultados de la aplicación del cuestionario.

Para la interpretación de los resultados se sugiere revisar los niveles de madurez para cada una de las temáticas propuestas. Los niveles de madurez serán utilizados como un medio para evaluar si las prácticas de la Universidad se alinean a la normativa vigente con relación a la gestión ética de la información.

Los resultados de la aplicación del cuestionario podrán ser analizados con base a la siguiente herramienta.

Tabla 4 Niveles de madurez para la gestión ética de los datos desagregados por componentes

GOBERNANZA INSTITUCIONAL Y SUPERVISIÓN INTERNA

I	La organización no tiene una estrategia, política o principios que incluyan la recopilación, el uso y el intercambio éticos de datos.
R	Los proyectos o disciplinas individuales consideran la ética de los datos para fines específicos. No existe una estrategia coherente para la ética de los datos en toda la organización.
D	Existe una responsabilidad clara y apoyo de la alta dirección para la consideración de la ética de los datos. La organización tiene roles que influyen en las decisiones sobre la ética de los datos.
G	La identificación y gestión de riesgos en torno a la recopilación, uso e intercambio de datos éticos es una parte fundamental del gobierno de la organización. También existen mecanismos para seguimiento y evaluación.
O	La organización tiene una clara estrategia, política y principios éticos relacionados con el uso y el intercambio de datos. La organización tiene un comité para hacer frente a las decisiones sobre ética.
HABILIDADES Y CONOCIMIENTOS	
I	La organización no proporciona apoyo y capacitaciones al personal. No hay claridad frente a la importancia de la gestión ética de los datos.
R	No existe una visión compartida frente a la gestión ética de los datos. La capacitación en temas de ética se enmarca en proyectos específicos.
D	El personal tiene una comprensión básica de la normativa aplicada a la gestión ética de los datos. Pares de trabajo hacen intercambio de conocimiento para la orientación de la gestión ética de los datos.
G	La organización tiene planes de capacitaciones en torno al tema de ética de los datos. La organización construye y fomenta redes de expertos dentro de la organización.
O	Todo el personal en los distintos niveles de la organización tiene la comprensión requerida para la gestión ética. La organización utiliza su experiencia y conocimiento del personal para proporcionar liderazgo en el área.
PROCESOS DE GESTIÓN DE RIESGOS	
I	No hay procesos definidos para gestionar los riesgos en la organización. Materialización de riesgos asociados al manejo poco ético de la información.
R	El tratamiento de riesgos no se da de forma articulada y transversal en la organización, sino que se ejecuta para fines muy concretos.
D	Los riesgos se priorizan en función del impacto esperado y se tratan en orden de prioridad. La organización tiene estandarizados los procesos para la gestión de riesgos frente al manejo ético de la información.
G	Todos los proyectos y disciplinas siguen procesos para identificar y evaluar el riesgo de daño al recopilar, usar y compartir datos. La organización implementa procesos de mitigación para los riesgos que tienen una mayor probabilidad de ocurrir.
O	La organización monitorea y se adapta a los cambios de las normas que influyen en el manejo ético para la recopilación, uso y uso compartido de datos e información. La organización implementa evaluaciones continuas para mitigar la materialización del riesgo.
FINANCIAMIENTO Y ADQUISICIONES	

I	No hay suficientes recursos para apoyar el uso e intercambio ético de los datos. El riesgo de que proveedores realicen prácticas no éticas en torno a los datos no se registra ni se mitiga.
R	La organización proyecta con claridad la dimensión ética de los datos en los proyectos y en las adquisiciones. Las consideraciones sobre ética de los datos se aplican en el marco de proyectos concretos, pero no existe un eje transversal y rutinario para aplicar estos principios en la organización.
D	La financiación de proyectos y los costos operativos incluyen habitualmente costos a largo plazo para evaluar, construir y demostrar la recopilación, el uso y el intercambio de datos éticos.
G	La organización tiene claridad sobre los costos/beneficios de las prácticas éticas de datos. La organización incluye una redacción estándar en todos los contratos que cubre las prácticas éticas de datos.
O	La organización lleva a cabo de manera proactiva la debida diligencia/auditoría en torno a las prácticas éticas de datos de las dependencias que utilizan, comparten y/o recopilan datos.
COMPROMISO DE LAS PARTES INTERESADAS Y EL PERSONAL	
I	Las comunicaciones sobre prácticas éticas de datos son densas, técnicas e ilegibles.
R	No existe una visión compartida frente a la gestión ética de los datos. La capacitación en temas de ética se enmarca en proyectos específicos.
D	Las comunicaciones sobre prácticas éticas de datos no usan el lenguaje de la comunidad reflejada en los datos o afectada por ellos (por ejemplo, usan lenguaje técnico interno).
G	La organización ha documentado un enfoque repetible para involucrar a las personas en la recopilación, el uso y el intercambio de datos, pero esto no se adopta ampliamente.
O	Es una rutina para todos los equipos involucrar a las comunidades reflejadas o impactadas por los datos y actuar en base a sus comentarios. La organización busca proactivamente comentarios de los clientes y el personal sobre cuánto confían en la organización con datos sobre ellos.
SITUACIÓN LEGAL Y CUMPLIMIENTO	
I	No hay conocimiento de los requisitos legislativos o de cumplimiento para proteger a las personas y las comunidades de cualquier daño.
R	El conocimiento de los requisitos de cumplimiento normativo y legal para proteger a las personas de daños se limita a unas pocas personas.
D	Existe una guía organizacional sobre el cumplimiento de los requisitos legales y sociales para evitar daños a las personas y las comunidades, pero esto no se adopta ampliamente.
G	La organización utiliza una estructura estandarizada para comprobar el cumplimiento de la legislación y las normas sociales pertinentes y esto se mide como parte de los informes de rendimiento.
O	La aplicación de las obligaciones legales de la organización en torno a la recopilación, el uso y el intercambio de datos se evalúan y actualizan de forma rutinaria para reflejar los requisitos en evolución y las prioridades sociales y políticas.

El uso del modelo de madurez de ética de datos es útil para identificar las falencias de las dependencias en torno al cumplimiento normativo. Para aprovechar la herramienta es importante que además de la

aplicación de los cuestionarios a los actores estratégicos se haga una revisión de los documentos de política, los procesos estandarizados y los roles asignados a la temática.

Posterior a tener claridad sobre el nivel de madurez de las prácticas éticas en la Universidad, es necesario que las dependencias comuniquen sus debilidades y fortalezas a las partes interesadas y los tomadores de decisiones con el fin de establecer un curso de acción frente las debilidades identificadas.

1.3.3.2. Socialización de los principios éticos para el manejo de los datos y la información

El objetivo de esta actividad es brindar orientación procedural para la elaboración de un plan de comunicación de los principios éticos en el manejo de los datos y la información, que contribuya a mejorar la comunicación estratégica de la Universidad, para lo cual también se requiere el compromiso de los tomadores de decisiones para impulsar su implementación.

Es indispensable planificar las acciones de comunicación y socialización en la medida que estas aportan a:

- Promover una visión compartida en torno al manejo ético de los datos en la organización.
- Fortalecer la imagen institucional en la medida que promueve que la comunidad universitaria actúe guiada por los principios éticos.
- Crea flujos de comunicación entre las distintas dependencias encargadas de la gestión de información.

Paso 1. Definir el objetivo del plan de comunicación.

El primer paso es definir los objetivos del plan de comunicación, estos pueden ser categorizados en objetivos estratégicos de largo plazo y objetivos tácticos de corto plazo. Para la redacción de los objetivos se puede tener en cuenta el diagnóstico del estado de las prácticas éticas de gestión de información, con el fin de actuar con base a los problemas y oportunidades identificados en este análisis.

Teniendo en cuenta que a partir de los objetivos se planifica la hoja de ruta de comunicaciones, es preciso que se establezcan no más de cinco objetivos que permitan atender las brechas y priorizar las acciones. Sumado a lo anterior, es importante que los objetivos tengan una definición con las siguientes características: específicos, cuantificables, reales, alcanzables y medibles.

Paso 2. Identificar el público objetivo

En el proceso de planificación del plan de comunicaciones es necesario realizar un análisis del público a quien va dirigido la comunicación. Eso implica conocer las necesidades particulares. Para esto es importante preguntarse

- ¿A qué tipo de población va dirigido el plan de comunicaciones?

- ¿Cuáles son los roles y responsabilidades de las audiencias a las que va dirigida el plan de comunicaciones?
- ¿Cuáles son las necesidades de la información particulares?
- ¿Cuáles son los mejores canales de comunicación para llegar a esa particular población?

Paso 3. Identificar los canales de comunicación

Para la ejecución de las acciones de comunicación es importante tener en cuenta los canales de comunicación disponibles, estos son administrados por UNIMEDIOS como dependencia encargada de producir y difundir la información investigativa, cultural, informativa y tecnológica de la Universidad. Algunos canales de comunicación que operan en el nivel nacional son:

- UN periódico
- Agencia de Noticias U.N.
- U.N Televisión
- U.N Radio
- Circular UN

Así mismo, el área corporativa de UNIMEDIOS es la encargada de poner en marcha estrategias para sensibilizar y motivar a la comunidad universitaria, en ese sentido, la dependencia deberá generar acciones en articulación con UNIMEDIOS con el fin de tener mayor visibilidad y posicionamiento.

Paso 4. Definir las acciones de comunicación

Teniendo claridad sobre los pasos anteriores, la dependencia puede iniciar el diseño del plan de acción, en el cual se definirán cada una de las acciones que se espera ejecutar en alineación con los objetivos propuestos, así como las herramientas de comunicación requeridas para su cumplimiento.

Algunas preguntas que son pertinentes para el diseño del plan de acción son:

- ¿Cómo se alcanzará la audiencia seleccionada?
- ¿Cuáles son los canales y medios de comunicación que se van a utilizar para difundir los principios éticos?
- ¿Qué herramientas se utilizarán para conseguir las metas?
 - Mensajes de correo electrónico
 - Publicaciones segmentadas
 - Cartas y circulares
 - Foros
 - Folletos

Paso 5. Definir el cronograma y asignar presupuesto

Posterior a tener claridad de las acciones que se van a ejecutar, es preciso establecer una programación con calendario de cada una de las acciones. Así mismo, se deberá asignar el presupuesto requerido para poner en marcha el plan de comunicación.

Paso 6. Diseñar herramienta para el control y seguimiento

El diseño del plan de comunicaciones deberá contemplar la construcción de una herramienta de seguimiento y control que permita medir el efecto de las acciones previstas. Para esto es importante construir una serie de indicadores de control y de resultados que se aplicaran para realizar mediciones con una periodicidad definida. Las actividades de seguimiento y control serán la base de la información para evaluar si se ha dado un cumplimiento efectivo de los objetivos planteados.

1.3.3.3. Formulación de metodología para la identificación de factores de riesgo éticos frente al manejo de los datos y la información.

Un componente muy importante del proceso de gestión ética de los datos es la implementación del marco general de gestión de riesgos, para lo cual, se hace necesario identificar aquellos factores de riesgos asociados al manejo ético de los datos.

En la actualidad la Universidad cuenta con un marco metodológico para la gestión de riesgo, este marco integra múltiples metodologías, y se basa en las directrices de la norma internacional de gestión del riesgo ISO 31000:2009. El documento público que presenta la guía para la implementación de riesgos de la Universidad aclara que, si bien es un avance significativo, es necesario que para versiones posteriores se incluyan riesgos de tipo estratégico y de seguridad de la información.

“Cabe resaltar que esta guía se está modificando constantemente, y, aunque el proceso de gestión de riesgos aplica para cualquier tipo, a la fecha de su publicación, solo contiene las pautas metodológicas y lineamientos particulares establecidos por la Universidad Nacional para gestionar los riesgos operativos y de corrupción; se espera incluir en próximas versiones otros tipos de riesgos como son los estratégicos y de seguridad de la información.” (UNAL, 2019, p. 8).

Teniendo en cuenta que la Universidad es una organización que gestiona grandes cantidades de información, es prioritario que la universidad avance en la identificación de los factores de riesgos asociados al uso e intercambio de información, así como en la definición de controles. Teniendo en cuenta que la información puede ser mal utilizada y utilizada en contravía a los principios relacionados con ética en el manejo de la información.

En ese orden de ideas, las dependencias encargadas de la gestión de la información pueden tomar en consideración los pasos propuestos a continuación para la identificación de los factores de riesgos.

Paso 1. Revisar el Marco Integral de la Gestión de Riesgos.

El cual tiene como objetivo establecer los lineamientos y las disposiciones para la articulación y gestión de riesgos a cargo de los procesos, las estrategias, los proyectos y los sistemas de gestión de la Universidad Nacional de Colombia, contribuyendo al cumplimiento de las estrategias y los objetivos misionales, a la reducción de eventos no deseados con sus impactos, a potenciar las oportunidades, la promoción de la cultura de la gestión del riesgo y a la creación y protección del valor.

Lo anterior, adaptando al contexto institucional referentes nacionales e internacionales que permitan diseñar, implementar, evaluar, mejorar, integrar y liderar un Marco unificado de políticas, responsabilidades, pautas metodológicas y tipologías de riesgos con el fin de identificar, analizar, valorar, tratar, monitorear, comunicar e informar de manera efectiva los riesgos a los que se expone la UNAL en todos sus niveles de aplicación con base en el apetito, la tolerancia y la capacidad definida.

Cabe mencionar que dicho documento institucional tiene cobertura en la totalidad de procesos, estrategias, proyectos y sistemas de gestión vigentes, al igual que cobija a todos los miembros y dependencias de la UNAL con responsabilidades frente a la gestión de las tipologías de riesgo.

Para la presente actividad, se abordará la Metodología para la gestión integral de riesgos en donde se definen las pautas metodológicas para la gestión integral del riesgo en la UNAL, tomando como referente y estructura las etapas del proceso de gestión de riesgos definidas en la Norma Técnica Colombiana ISO 31000:2018.

- **Etapa 1. Establecimiento del contexto:** se relaciona con la claridad del entorno en el cual se gestionarán los riesgos, los procesos que involucra, el flujo de los procesos y los objetivos. Este se realiza desde diferentes niveles: Gestión integral del riesgo, Estratégico, Sistemas de Gestión, Procesos, Proyectos, otros.
- **Etapa 2. Identificación del riesgo:** para esta etapa es necesario identificar aquellas situaciones que obstaculizan el cumplimiento de los objetivos. Por su parte el DAMMA establece que los riesgos éticos deben relacionarse con los principios, es decir, eventos que pueden suceder si no se cumplen los principios.

Figura 2 Proceso para la gestión del riesgo de la UNAL



Así pues, la **Figura 2** presenta el conjunto de componentes que guían el accionar al interior de la universidad en materia de monitoreo de riesgos.

La etapa de identificación de riesgos es central para el desarrollo de esta actividad. Siguiendo la guía de la Universidad, es muy importante seguir el paso a paso para la redacción de riesgos. Para lo cual debe tenerse en cuenta:

- El activo de la información amenazados: recursos tangibles o intangibles que están susceptibles a ser afectados por una amenaza.
- La amenaza: Circunstancias con efectos negativos sobre los activos de información.

Además de lo anterior, es preciso que se tengan en cuenta los siguientes lineamientos propuestos en la Guía de Gestión de Riesgos de la Universidad al momento de redactar los riesgos identificados (UNAL, 2019, p. 56).

- Evitar la negación para expresar el riesgo.
- Tomar como punto de partida las debilidades y/o amenazas encontradas en la formulación del contexto del proceso.
- No se debe confundir con un problema.
- No se debe redactar en términos de una no conformidad o incumplimiento.
- Debe existir interrelación entre la amenaza (condición o situación) y el activo amenazado (Recurso, sujeto, objeto).

Paso 2. Identificar marcos de referencia y buenas prácticas en la gestión de riesgos de información.

Además del marco de referencia propuesto por el SIGA para la gestión de riesgos, es preciso que los responsables de la identificación de riesgos tengan presente otras experiencias de buenas prácticas.

Algunas prácticas que pueden ser tenidas en cuenta para la identificación de riesgos son:

- Principios para resguardar la privacidad y flujos transfronterizos propuestos por la Organización para la Cooperación y el Desarrollo Económico.
- Reglamentación de los ficheros computarizados de datos personales propuesta en la Asamblea General de las Naciones Unidas el 14 de diciembre de 1990.
- Reglamento General de Protección de Datos (RGPD)

1.3.4. Salidas

Diagnóstico de cumplimiento normativo: Documento que evidencie el resultado de la implementación de la metodología de análisis de brechas y la medición de las prácticas de la Universidad en torno al manejo ético de la información en concordancia con el modelo de madurez propuesto.

Plan de comunicación de los principios éticos: Documento técnico de planificación con las acciones, responsables y herramientas para socializar los principios éticos.

Documento con la identificación de factores de riesgo en la gestión ética de los datos: Documento con el listado de riesgos asociados a la gestión ética de los datos.

1.3.5. Usuarios

Comunidad Universitaria: Está conformada por estudiantes matriculados en los diferentes programas curriculares de la Universidad Nacional de Colombia, el personal docente e investigador, el personal Administrativo y de Prestación de servicios y egresados.

Líderes de los sistemas de información: personas que lideran los sistemas de la información y, por tanto, son quienes promueven la aplicación de prácticas adecuadas en la ejecución de sus labores cotidianas.

Comité de Gobierno de la información: Instancia principal de articulación interinstitucional encargada de los diseños programáticos y metodológicos en temas de gestión de la información de la Universidad Nacional de Colombia.

2. GOBIERNO DE DATOS

2.1. Conceptualización

Como entidad líder en la educación e investigación en el país, la Universidad Nacional de Colombia gestiona en su día a día gran cantidad de datos que se convierten en un activo de gran relevancia para la institución y cuyo manejo es un desafío que exige de la administración buscar las mejores estrategias que le permitan realizar una adecuada gestión de sus datos. Esto requiere de planificación, almacenamiento, control, interoperación, disposición final y cada una de las estrategias que permitan que los datos sean una fuente suficiente de la información para la toma de decisiones basada en evidencia.

Para lograr la adecuada gestión de los datos es fundamental que exista un componente central y transversal al manejo de los datos, este se identifica como gobierno de los datos. De acuerdo con la metodología del DAMA, el gobierno de datos se define como “el ejercicio de autoridad y control (planificación, seguimiento y ejecución) sobre la gestión de los activos de los datos” (DAMA Internacional, 2017, p. 69). El Ministerio de Tecnologías de la Información define el gobierno del dato como la

“... Disciplina clave para controlar el uso de los datos maestros del sector público, además de abordar con éxito las renovaciones, migraciones en sistemas y organizaciones asociadas con el dato. El gobierno aborda los ámbitos de arquitectura, calidad, custodia, aprovisionamiento y gestión de la demanda del dato” (Ministerio de Tecnologías de la Información y las Comunicaciones, 2014, p. 48)

Considerando las definiciones anteriormente expuestas, la Universidad deberá centrar sus esfuerzos en la generación de un gobierno del dato que tome la actual estructura con que se cuenta para su administración y la fortaleza desde las dimensiones de Modelado y Diseño de Datos, Gestión de Datos, Maestros y Referencias, Integración e Interoperabilidad, Arquitectura, Ética y Calidad de Datos, con el fin de estructurar un gobierno del dato que le permita planificar, hacer seguimiento y controlar la gestión de sus datos como un activo valioso para la institución.

Si consideramos nuevamente el modelo propuesto por el DAMA, el gobierno de datos “se centra en cómo se toman las decisiones sobre los datos y cómo se espera que se comporten las personas y los procesos en relación con los datos. El alcance y el enfoque de un programa particular de gobierno de datos dependerá de las necesidades de la organización” (2017, p. 70) (DAMA, 70). Muchos de estos programas incluyen: Estrategia, Política, Normas y Calidad, Supervisión, Cumplimiento, Gestión de Problemas, Proyectos de Gestión de Datos y Valoración de activos de datos.

De acuerdo con las definiciones de DATAVERSITY, la gobernanza de datos busca que se cumplan las políticas que ya existen dentro de la organización. Según esta definición, los elementos que una política de gestión de datos debe contemplar son: seguridad, privacidad y uso de datos, normas de seguimiento, intercambio de datos, retención y vencimiento de datos, calidad de datos, integración de datos,

actividades de desarrollo, funciones y responsabilidades, inteligencia de negocio y cumplimiento y normativa.

Lo anterior permite inferir que, aun cuando la Universidad no cuenta con una estructura claramente definida sobre el gobierno de los datos, los ha venido gestionando de manera implícita en sus procesos. Dicha estructura debe revisarse para iniciar -a partir de ella- una construcción formal de un gobierno de datos con una visión a futuro desde la generación de una política que le permita desarrollar posteriormente los otros componentes y en concordancia con los elementos comunes que se pueden identificar en los modelos propuestos por DAMA, DATAVERSITY e ISO.

Igualmente debe tenerse en cuenta que el gobierno de datos dentro de la Universidad debe pasar de una estrategia temporal a ser parte del ADN de la gestión y de los procesos con el fin de darle continuidad y mejora continua que permita generar valor y apoyar la toma de decisiones basada en evidencias por parte de la alta dirección. En este sentido, es necesario recordar que “El gobierno de datos no es algo que se hace una vez. La gobernabilidad de los datos requiere un programa continuo enfocado en asegurar que una organización obtenga valor de sus datos y reduzca el riesgo relacionado con sus datos” (DAMA Internacional, 2017, p. 73).

También cabe mencionar que las recomendaciones del MinTIC, la Misión de Sabios 2019, las consideraciones de la OCDE en materia del avance hacia lo digital y de conformidad con el numeral 2 del artículo 15c del Acuerdo 316 de 2019 del Consejo Universitario, el plan institucional en materia de tecnologías de la información de la Universidad Nacional de Colombia deben tener un enfoque de generación de valor público que habilite las capacidades y servicios de tecnología necesarios para impulsar la transformación digital de la Universidad y su eficiencia y transparencia. Así mismo, la dirección universitaria debe establecer los ajustes organizacionales necesarios para garantizar un adecuado gobierno de la información.

2.1.1. Modelo de gobierno de datos

Definir una política de gestión de datos y las dimensiones que la componen es tan importante como definir el tipo operativo de gobierno de datos que se tendrá en la institución y que concuerde con otras políticas y la visión a futuro que ella tenga. Al respecto, DAMA establece 3 maneras de marcos operativos de gobierno de datos (**Figura 3**): un modelo centralizado, un modelo replicado y un modelo federado. En el modelo centralizado, una organización de gobierno de datos supervisa las actividades en todas las áreas temáticas. En el modelo replicado, cada unidad de negocio adopta el mismo modelo operativo y estándares del Gobierno de Datos. En el modelo federado, una organización de gobierno de datos coordina con varias unidades de negocio para mantener definiciones y estándares coherentes (DAMA Internacional, 2017, p. 77)

Figura 3 Marcos operativos de gobierno de datos

Centralizado



Federado



Replicado



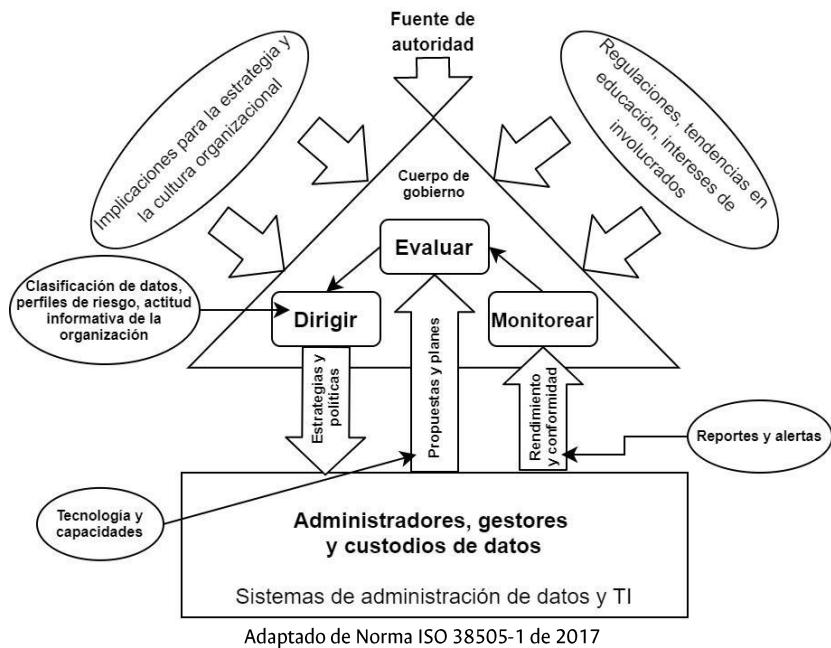
Adaptado de DAMA (2017, p. 77)

Un cuerpo de gobierno tiene tres tareas fundamentales (localizadas en el triángulo central de la **Figura 4**): i) evaluar el uso de datos (presente y futuro); ii) dirigir la implementación de las políticas y las estrategias de uso de datos para dar cumplimiento a los objetivos de la organización; iii) verificar el cumplimiento de las políticas y sus resultados respecto de las estrategias. Los demás aspectos de la gestión de datos deben asignarse a los administradores dentro de la organización y así, el gobierno central y los mecanismos de apoyo conseguirán hacer un uso efectivo, eficiente y aceptable de los datos.

Sin embargo, la organización a nivel general tiene otros requerimientos que debe satisfacer (**Figura 4**), estos son: la cultura de uso de datos que la rodea y la toma de decisiones que estén de conformidad con ella y con los objetivos organizacionales; las presiones externas del mercado, los clientes, los competidores y las normas y regulaciones propias del negocio; los procesos de evaluación de los dos puntos anteriores además del uso de datos y las capacidades de la organización. Además, la dirección, es decir, la asignación de responsabilidades para maximizar el valor de los datos, controlar los riesgos

asociados y garantizar un nivel apropiado de la gestión de datos y finalmente, el monitoreo de los resultados de la organización frente al uso de datos.

Figura 4 Modelo de gobierno de los datos



Adaptado de Norma ISO 38505-1 de 2017

2.1.2. Estado de la gestión de los datos y alineación con la misionalidad de la Universidad Nacional de Colombia

El seguimiento del estado de la gestión de los datos de una organización responde a la necesidad de llevar a cabo un proceso de evaluación constante como parte de un programa de gobierno de datos. Esto se hace con el propósito de establecer de forma oportuna la eficiencia, los beneficios y los riesgos de las políticas y las directrices que la organización desarrolla con relación al uso de datos. A partir de ahí, la organización va consolidando insumos que más tarde le permitirán mejorar la calidad y el uso que le da a los datos que maneja. Cabe resaltar que

La evaluación de la calidad de los datos proporcionará una visión de los problemas y obstáculos existentes, así como el impacto y el riesgo asociado con los datos de baja calidad. La evaluación de la calidad de los datos puede identificar los procesos de negocio que están en riesgo si utilizan datos de baja calidad, así como los beneficios financieros y otros beneficios de crear un programa de calidad de datos como parte de los esfuerzos de Gobierno de Datos (DAMA Internacional, 2017, p. 82).

Lo anterior es importante porque los datos representan un activo crucial para cualquier organización, especialmente en la actualidad con el avance de las tecnologías y la importancia que va adquiriendo la información en distintos mercados globales. Esto ha hecho que cada vez sea más provechoso económicamente procesar grandes grupos de datos y, en muchos casos, que esta se vuelva la principal área de negocio de una organización independientemente de cuál sea su ocupación principal. El potencial de la información es bastante alto y puede generarle un nuevo valor a la organización complementando vacíos, proporcionando datos adicionales o sirviendo como insumo para la toma de decisiones.

Por supuesto, con el incremento de la cantidad y la variedad de datos que se gestionan, las obligaciones y los riesgos también se multiplican; por eso es importante tener una adecuada planeación que permita responder óptimamente a las nuevas necesidades. En cuanto a los riesgos, es fundamental que el cuerpo de gobierno tenga establecido cuál es el umbral permitido y sepa controlarlo, sobre todo, en la medida en que los datos se hacen “estratégica, operacional y financieramente importantes para la organización” (ISO & IEC, 2017, p. 16).

Entre los riesgos más altos se encuentran, por un lado, calidad, ya que tener datos de baja calidad hace que los procesos del negocio se vean afectados y, por otro lado, el riesgo de desaprovechar los datos con que cuenta una organización en lugar de darles utilidad, porque podrían impactarse las operaciones de la organización, la seguridad de los datos e incluso provocarse riesgos financieros (ISO & IEC, 2017). En este sentido es importante saber administrar los roles y las obligaciones que deben cumplirse. Esto implica que según las necesidades de la organización el cuerpo de gobierno de datos puede formar subcomités que estarán encargados de tareas específicas, con la intención de ir consolidando la estrategia con la cual se van a gestionar los datos (ISO & IEC, 2017).

Si bien se cuenta con ciertos mecanismos de apoyo, la responsabilidad de que se haga una buena gestión de datos siempre recae sobre el cuerpo de gobierno; esto quiere decir que tiene tareas que están vinculadas al funcionamiento de la organización y no pueden ser evadidas o aplazadas. La primera de ellas es reconocer la relevancia que los datos tienen dentro de la misión de la organización; la segunda, es asegurarse de que todos los miembros de la organización también conozcan ese valor; la tercera es garantizar que exista una estructura adecuada para el gobierno y la gestión de datos; en cuarto lugar, hacer una evaluación permanente del mecanismo de gobierno y de la gestión de datos, especialmente a través de procesos externos de análisis y verificación.

Para esta labor la norma ISO 38505-1 propone una serie de áreas y obligaciones del gobierno de datos que en otros términos son etapas de un proceso. Estas sirven como marco para hacer la “evaluación, monitoreo y dirección de las actividades organizacionales para el gobierno de datos en general” (ISO & IEC, 2017, p. 16) y así determinar cuáles son los controles pertinentes, cómo se deben modificar las estrategias, las directrices, etc.

Este es un trabajo que se realiza de forma permanente, es cíclico, puesto que mientras las operaciones de la organización siguen su curso está se va a ir transformando paulatinamente en distintas áreas, como la cantidad y la clase de datos que produce, el valor que les da y las limitaciones que estos suponen, así como las transformaciones externas, del mercado, jurídicas, entre otras.

2.1.3. Estrategia de gobierno de datos

El cuerpo de gobierno de los datos es responsable de que la estrategia de datos esté alineada con la planeación estratégica general de la organización, incluyendo con sus actuales y futuras capacidades. La estrategia de gobierno de los datos plantea la visión de gobierno en una organización y opera de conformidad con la estrategia general de la organización, incidiendo en cada uno de los procesos de la gestión de datos y TI.

Con la estrategia se explicitan las metas, principios y políticas que regirán el desempeño de la organización y son el medio que emplea el cuerpo de gobierno para conseguir un uso de datos que sea “efectivo, eficiente y aceptable” (DAMA Internacional, 2017; ISO & IEC, 2017, p. 5). Además, es un apoyo que enruta los esfuerzos de la organización para que satisfaga sus obligaciones, que van desde las demandas de los agentes externos y la observancia de la normativa que regula el negocio, hasta la garantía de la calidad de los datos en términos de protección, privacidad y (ISO & IEC, 2017). Todo lo anterior sin dejar de lado los objetivos propios del gobierno que buscan sacar el máximo beneficio posible de la inversión que se hace en los datos, sabiendo responder a los riesgos que se van presentando, teniendo en cuenta los obstáculos que se puedan presentar, y haciéndolo con diligencia y precisión.

Hay seis principios que todo gobierno de TI debería introducir en el desempeño de los distintos niveles de su estrategia de gobierno de datos, que son **i) la responsabilidad, ii) la estrategia, iii) la adquisición, iv) el rendimiento, v) la conformidad y vi) el desempeño humano**. Tales principios, a su turno, orientan los grandes deberes del gobierno: dirigir, monitorear, evaluar (ISO & IEC, 2017).

Esos principios se asignan posteriormente para que puedan gestionarse de forma particular, convirtiéndose en aspectos específicos del gobierno relativos a: el valor (beneficio que se extrae de los datos brutos en calidad, cantidad, temporalidad y contexto para generar conocimiento útil); el riesgo (en todos los niveles) y las limitaciones (del uso, normativas, contractuales, éticas, sociales, organizativas).

Los niveles de gobierno son las instancias entre las que se dividen las tareas administrativas, técnicas o de cualquier tipo. Para saber qué tipos de asignaciones son necesarias, de cuántas se trata y cómo deben ejecutar sus funciones, es necesario construir y poner en práctica un modelo operativo. Este brindará nociones indispensables como:

- El valor que tienen los datos para la organización (si son principales o secundarios).
- El modelo de negocio, en otras palabras, los factores que lo producen (ubicación, estructura, etc.).
- Los factores culturales (flexibilidad al cambio, principios, vías de resolución de problemas, etc.).
- El impacto de la regulación (nivel de incidencia de las normativas) (DAMA Internacional, 2017).

2.1.4. Roles y responsabilidades en el gobierno de datos

Para el desarrollo de un marco de gobierno de datos es necesario definir el papel que juegan los diferentes actores. En primer lugar, el marco de referencia del DAMA propone tipos de custodios de los

datos, quienes tienen una relación directa con la estructura organizacional, sus diferentes niveles de desagregación de funciones y las dinámicas propias del sector. Entre los tipos de custodios están: jefe ejecutivo, empresarial, de negocio, dueños de datos y técnicos coordinadores. En segundo lugar, DAMA describe unos órganos típicos de gobierno, que presentamos en la **Tabla 5**.

Tabla 5 Órganos de gobierno. Adaptado de: DAMA (2017, p. 76)

Órganos de gobierno de datos	Certificación
Comité Directivo de Gobierno de Datos	Es un organismo conformado por un grupo multifuncional de altos ejecutivos con la mayor autoridad para el gobierno de datos en una organización. Este comité es el responsable de la supervisión, el apoyo y la financiación de las actividades de gobierno de datos. Por lo general libera fondos las actividades de acuerdo con lo recomendado por el Consejo de Gobierno de Datos y el Oficial en jefe de Datos (<i>Chief Data Officer</i>). Este comité puede, a su vez, supervisar fondos de alto nivel o comités directivos basados en iniciativas.
Consejo de Gobierno de Datos	Gestiona las iniciativas (por ejemplo, el desarrollo de políticas o métricas), problemas y escalamientos de gobierno de datos. Está conformado por ejecutivos según el modelo operativo utilizado.
Oficina de Gobierno de Datos	Enfoque continuo en las definiciones de datos a nivel empresarial y los estándares de gestión de datos en todas las áreas de conocimiento del DAMA. Consiste en coordinar roles que están etiquetados como custodios y dueños de datos (<i>data stewards</i>).
Equipo de Custodia (<i>stewardship</i>) de datos	Son comunidades de interés centradas en una o más áreas temáticas o proyectos específicos. Colaboran o consultan con equipos de proyectos sobre definiciones de datos y estándares de gestión de datos. Está conformado por analistas de datos y <i>data stewards</i> técnicos y de negocio.
Comité Local de Gobierno de Datos	El Comité Local de Gobierno de Datos es una alternativa para que organizaciones pequeñas eviten la complejidad de tener Consejos de Gobierno de Datos divisionales o departamentales que trabajan bajo los auspicios de un Consejo de Gobierno de Datos empresarial.

2.1.5. Gestión del cambio

De acuerdo con DAMA, “La Gestión del Cambio Organizacional (OCM) es el vehículo para lograr cambios en los procesos y sistemas de una organización” (2017, p. 86). Esta afirmación es clave entendiendo que todo cambio a nivel organizacional de tecnologías, de procesos, normativo; entre otros, debe ir acompañado de una serie de actividades que permitan la interiorización de las nuevas formas para el

desarrollo de las tareas en los colaboradores y así se logren los objetivos propuestos al incorporar estos nuevos elementos.

El marco de referencia del DAMA propone la conformación de un equipo responsable de:

- Planeación, capacitación y comunicaciones.
- Implementación de políticas.
- Promoción del valor de los activos de datos.
- Implementación de capacitaciones en gestión de datos.
- Implementación de nuevas métricas y KPIs.
- Influir en el desarrollo del sistema.
- Monitoreo y retroalimentación de las actividades de Gobierno de Datos.

A nivel interno, la Universidad ha venido adoptando la Guía de Gestión del Cambio U.GU.SIGA.002 (2020) como ejercicio de mejora en sus sistemas de gestión y como instrumento de referencia para la gestión de los cambios a nivel institucional. El objetivo de esta guía es establecer la metodología para “la atención oportuna de la gestión del cambio dentro de la Universidad por medio de su identificación, diagnóstico, planeación y seguimiento, a fin de generar estrategias necesarias que permitan mantener la integridad y mejora continua de la gestión institucional” (Universidad Nacional de Colombia - SIGA, 2020, p. 2).

Se espera que la Guía de Gestión del Cambio sea utilizada para todo tipo de cambio en la Universidad (sea este estratégico, de procesos, infraestructura, estructura organizacional, tecnología, cultura organizacional, gestión humana o normativo), coadyuvando a la incorporación de nuevas dinámicas a la gestión universitaria, mitigando los impactos negativos y potenciando los beneficios relacionados con el cambio implementado. Para que esto sea posible, se establecieron seis pasos para aplicar la guía:

1. Identificar el cambio.
2. Identificar los gestores del cambio.
3. Definir el plan de acción de la gestión del cambio.
4. Ejecutar los planes de acción para la gestión del cambio y evaluar.
5. Hacer seguimiento y evaluar la gestión del cambio.
6. Mejorar continuamente.

Definir y desarrollar una metodología que permita realizar una adecuada gestión del cambio permite a los colaboradores adaptarse mejor a los cambios, asimilarlos positivamente y evita la generación de obstáculos, movilizando todas las capacidades hacia una mejora continua.

2.2. Propósito

Establecer un modelo de gobierno de datos e información que se articule con el modelo de gestión de la Universidad Nacional. Para ello es necesario determinar actores, roles y responsabilidades que permitan la construcción de lineamientos, metodologías y herramientas para la dirección, seguimiento, monitoreo y evaluación de la gestión de datos. Todo lo anterior de cara a la toma de decisiones basada

en evidencia, asegurando su cumplimiento y determinando los criterios que permitan el mejoramiento continuo.

Este modelo está alineado con estándares internacionales y con la visión 2034 de la Universidad de acuerdo con el Plan Estratégico Institucional (PLEI). Además, concuerda con la etapa de transformación digital de la Universidad en su propósito superior de apoyar la formación de ciudadanos integrales como agentes de cambio con conciencia social, y apoyar su desarrollo en la ética, la gestión del cambio, el liderazgo colectivo, entre otros, a través de diversas estrategias y el uso de herramientas digitales, como elemento central para la aplicación del desarrollo de las funciones misionales de la comunidad universitaria. (Transformación Digital 2030 - Propuesta de una política institucional borrador - 2021, pág. 7)

Dentro de su modelo de gobierno de datos, la Universidad deberá iniciar la consideración del dato como un activo para empezar a gestionarlo como tal, esto con el fin de ajustarlos de manera coherente con el nuevo modelo, determinar con base en la actual estructura de la Universidad cuáles son los roles y responsables que intervienen en la gestión de los datos y para proponer a partir de ello la construcción de una estructura más robusta y acorde con la consecución de las metas esperadas.

Todo lo anterior redunda en que la gestión de los datos dentro de la institución sea reconocida como un activo y conduzca a facilitar la toma de decisiones basada en evidencia; asegurando así el cumplimiento y determinando los criterios de medición que permitan el mejoramiento continuo y el cumplimiento de los objetivos y metas de mediano plazo.

2.2.1. Metas

- Establecer un esquema de gobierno de los datos y la información que permita la orientación de la gestión en este aspecto y su articulación con el modelo de gobernanza de la Universidad.
- Definir lineamientos que permitan el seguimiento y monitoreo de la gestión de datos e información en la Universidad.
- Establecer pautas para la evaluación de la gestión de los datos y la información articulándose con los sistemas de gestión de la Universidad de cara al mejoramiento continuo.
- Fomentar el mejoramiento de la calidad de los datos y la información para la toma de decisiones.
- Definir roles y responsabilidades en la gestión de los datos y la información.
- Desarrollar una estrategia que permita la identificación de los datos como un activo institucional necesario para la toma de decisiones, como base para el desarrollo de proyectos en la materia.
- Incorporar un proceso constante de gestión del cambio que fortalezca la adopción de la gestión de los datos y la información y su importancia institucional.

2.2.2. Alcance Fase II

Para la Fase II de implementación del componente de gobierno del Modelo de Gestión de la Información, se debe partir por poner en funcionamiento el denominado Comité de Gobierno de la Información (lo cual no está en detrimento del desarrollo paralelo de la implementación o perfeccionamiento de los

demás componentes del Modelo de Gestión de la Información), el cual debe llevar a cabo las demás actividades establecidas en la dimensión. El Comité de Gobierno debe iniciar por definir su agenda y por priorizar la implementación de temas incluidos en la Política de Gestión de la Información de la UNAL. Luego de esto, deberá ajustar el presente documento de Modelo de Gestión de la Información en conformidad con la agenda y la priorización realizada, y posteriormente, deberá elaborar una Estrategia de Gobierno para toda la universidad en los términos establecidos en la sección de conceptualización.

Para las siguientes fases se propone que se desarrollen los temas asociados a las metodologías e instrumentos requeridos para el seguimiento y monitoreo de la gestión de los datos; la evaluación, de acuerdo con los indicadores de calidad que se establezcan; las técnicas para la valoración de los activos de datos; los criterios para la gestión de problemas y comunicaciones y los criterios para la viabilidad de proyectos de gestión de datos.

2.3. Desarrollo fase II dimensión de gobierno de datos

Tabla 6 SIPOC para el desarrollo de la dimensión de gobierno de datos

Proveedores	Entradas	Actividades	Salidas	Usuarios
DNED	Política de GI Modelo de GI PLEI	Organización del Comité de Gobierno de la Información	Acta de conformación del Comité de Gobierno de GI	Todos los actores del sistema de GI
DNED – Oficina de GI	Política de GI Modelo de GI PLEI PGD PETI	Definición de agenda de trabajo del Comité de Gobierno de la Información	Agenda de trabajo de Comité de Gobierno de GI	Comité de Gobierno de GI
DNED – Oficina de GI	Política de GI PLEI PGD	Priorización de temas incluidos en la Política de Gestión de la Información	Lista de temas priorizados	DNED – Oficina de GI
DNED – Oficina de GI	Modelo de GI PLEI PGD PETI	Revisión y ajuste del Modelo de Gestión de la Información (incluyendo instituciones orientadas al perfeccionamiento del gobierno de datos)	Modelo de GI Ajustado y revisado	Rectoría CNED
DNED – Oficina de GI	Lista de temas priorizados Modelo de GI ajustado y revisado	Formulación de Estrategia de Gobierno	Plan de trabajo GI	Todos los actores del sistema GI

2.3.1. Proveedores

En las primeras fases que se desarrollen en la Universidad para la implementación de la dimensión de gobierno el liderazgo de la DNED será el eje fundamental. Como puede observarse en el diagrama SIPOC, la DNED (particularmente la Oficina de Gestión de la Información) será el principal proveedor de las entradas necesarias para llevar a cabo las actividades establecidas. Esto se debe a que las entradas corresponden a productos que han sido diseñados, socializados y ajustados bajo el liderazgo de la

DNED, por lo que esta área ejercerá el rol de custodio de las últimas versiones de la Política de Gestión de la Información, el Modelo de Gestión de la Información (incluyendo sus anexos), y de las actas que se realicen como resultado de las reuniones del Comité de Gobierno de la Información.

2.3.2. Entradas

- Política de Gestión de la Información.
- Modelo de Gestión de la Información.
- PLEI - Plan Estratégico Institucional.
- PGD - Plan Global de Desarrollo.
- PETI - Plan Estratégico de Tecnologías de Información.
- Política de Tratamiento de Datos UNAL.
- Política de Seguridad Informática y de la Información.

2.3.3. Actividades

2.3.3.1. Organización del Comité de Gobierno de la Información

La Política de Gestión de la Información establece:

ARTÍCULO 8. Comité de Gobierno de la Información. Se establece el Comité de Gobierno de la Información como instancia de coordinación suscrita del Comité Nacional de Estrategia Digital, el cual será la instancia principal de articulación institucional en cuanto al tema de gestión de la información, y estará conformado por:

- Vicerrectoría General
- Dirección Nacional de Estrategia Digital o quien haga sus veces
- Dirección Nacional de Planeación y Estadística

PARÁGRAFO 1. La Secretaría Técnica del Comité de Gobierno de la Información será ejercida por parte de la Oficina de Gestión de la Información de la DNED, y tendrá a cargo todo el apoyo técnico y logístico para el adecuado funcionamiento del Comité de Gobierno de la Información.

PARÁGRAFO 2. Los miembros del Comité de Gobierno de la Información podrán invitar a otras dependencias que consideren pertinentes a sus sesiones, y trabajar de manera articulada con estas para el diseño e implementación de los lineamientos que se requieran.

El Comité de Gobierno de la información tendrá el rol de cuerpo de gobierno de los datos y la información en la Universidad. En este sentido, es necesario iniciar labores de socialización e impulso de la Política de Gestión de la Información, así como de las conceptualizaciones y lineamientos procedimentales y metodológicos que se incluyen en el Modelo de Gestión de la Información. Estas labores de socialización e impulso deben dirigirse a quienes serán los miembros permanentes y transitorios del Comité de Gobierno de la Información, quienes deberán definir un acta de conformación, y deliberar acerca de los diseños institucionales que la Oficina de Gestión de la información de la DNED

elabore acerca de roles, responsabilidades, objetivos, procedimientos, metodologías de seguimiento y evaluación, entre otros, dirigidos a la organización del Comité de Gobierno de la Información.

Desde la Oficina de Gestión de la información de la DNED, así como desde el Clúster de Gestión de la información se recomienda que los delegados por parte de las dependencias con roles permanentes y transitorios dentro del Comité de Gobierno de la Información adquieran la competencia de asistir de manera regular a las sesiones del Comité (una vez a la semana si es posible). Así mismo, se recomienda que las personas que se asignen como asistentes a estos espacios cuenten con suficiente conocimiento en el tema de gestión de información, y tengan capacidad de decisión sobre estos temas.

2.3.3.2. Definición de agenda de trabajo del Comité de Gobierno de la Información

La primera actividad que debe ser realizada por parte del Comité de Gobierno de la Información una vez esté conformado, es el establecimiento de un acuerdo alrededor de su agenda de trabajo a corto o mediano plazo. En lo preferible, esta agenda de trabajo debería estar alineada con la agenda organizacional en materia de gestión de información, la cual puede verse reflejada en el Plan Global de Desarrollo, el Plan Estratégico Institucional, el Plan Estratégico de Tecnologías de la Información, y finalmente, la Política y el Modelo de Gestión de la Información.

Teniendo en cuenta que el Comité de Gobierno de la Información corresponderá a un escenario de coordinación institucional, en el que se reunirán varias dependencias con distintos intereses y agendas, es necesario aclarar que esta agenda debe resultar de un proceso continuo de deliberaciones técnicas. Al ser la gestión de la información un tema complejo y con múltiples facetas, su desarrollo en la universidad puede realizarse de manera diversa y múltiple, por lo que es necesario realizar acuerdos a nivel institucional que garanticen una efectiva coordinación entre los principales actores implicados.

En las etapas tempranas de funcionamiento del Comité de Gobierno de Información, la agenda de trabajo puede dirigirse a discutir y perfeccionar los diseños institucionales orientados al mismo funcionamiento del Comité de Gobierno de la Información, así como de la institucionalidad de apoyo que le permitirá ejercer gobernabilidad y ser eficaz frente a los objetivos que se tracen. Tal como se mencionó en el punto anterior, estos diseños institucionales primarios estarán orientados a determinar roles, responsabilidades, objetivos, procedimientos y metodologías de seguimiento y evaluación del Comité de Gobierno de la Información, de tal manera que se determine al interior de la Universidad el modo en el que se desarrollarán las tres funciones de gobierno que han sido establecidas por la Norma ISO 38505-1 de 2017.

2.3.3.3. Priorización de temas incluidos en la Política de Gestión de la Información

Uno de los principales temas que deben incluirse dentro de la agenda de trabajo del Comité de Gobierno de la Información es lo relativo a la selección de temas que han sido asignados como de su competencia al interior de la Política de Gestión de la Información. Esta política, se formuló a través del Clúster de Gestión de Información, lo que garantizó una amplia participación para la construcción de un acuerdo

conjunto entre las distintas dependencias de la Universidad, por lo que garantizar la gobernabilidad para su implementación es fundamental para el desarrollo de la gestión de la información dentro de la organización.

En este acuerdo institucional, el Comité de Gobierno de la Información quedó encargado de liderar la implementación de los siguientes elementos:

1. Determinar las demás dependencias que participarán de manera transitoria en el Comité de Gobierno de la Información
2. Diseñar los roles y responsabilidades para la gestión del ciclo de vida del dato, y ponerlo en consideración de la Rectoría
3. Actualizar el Modelo de Gestión de la Información
4. Definir los criterios y lineamientos para la identificación regular de necesidades de información
5. Revisar y aprobar (previo diseño de la DNED) el plan de acción del Modelo de Gestión de la Información
6. Establecer (en compañía de la Rectoría) los componentes de la arquitectura de datos para la UNAL
7. Formular y establecer el diseño de flujo de datos, junto con lineamientos para el almacenamiento y procesamiento de bases de datos, aplicaciones, plataformas y redes
8. Proponer (para aprobación de la Rectoría) estándares para el lenguaje común y los datos
9. Articular diseños normativos y técnicos con estrategias que se diseñen en materia de almacenamiento y procesamiento de datos
10. Definir normas técnicas y estándares institucionales relacionados con la calidad de los datos y la información
11. Definir lineamientos para la seguridad de los datos y la información que deben aplicar custodios de datos
12. Definir el modelo de cálculo económico de los datos y la información de la Universidad
13. Definir lineamientos para la transformación e integración de datos no estructurados al interior del Modelo de Gestión de la Información
14. Evaluar la Política y el Modelo de Gestión de la Información

Como puede observarse, existen todas estas competencias del Comité de Gobierno de la información que son fundamentales para ejercer una adecuada gobernabilidad de la información y los datos en la Universidad, sin embargo, es responsabilidad del Comité el establecer un orden de priorización entre todos estos elementos. Esta decisión deberá tomarse con base en el balance que pueda realizarse con respecto del alcance de la gobernabilidad que logre tener en Comité, así como de la capacidad institucional que se pueda utilizar para desarrollar estas competencias. Adicionalmente, el Comité de Gobierno de la Información deberá establecer lineamientos acerca del mejor modo de implementar las competencias que se terminen priorizando, de tal manera que la DNED pueda ejercer su labor de secretaría técnica del Comité, con suficientes lineamientos y capacidad.

2.3.3.4. Revisión y ajuste del Modelo de Gestión de la Información (incluyendo instituciones orientadas al perfeccionamiento del gobierno de datos)

Teniendo presente la agenda de trabajo acordada, así como la priorización realizada frente a las competencias incluidas en la Política de Gestión de la Información, una de las principales responsabilidades que se deberá realizar en materia de gobierno de datos e información corresponde a la revisión y ajuste continuo del presente Modelo de Gestión de la Información. Anteriormente se ha manifestado como propósito de este modelo la orientación de la gestión de la información a través de normas, actores, herramientas, instrumentos y procesos, esto en correspondencia con el supuesto de que los marcos de gestión de datos se usan para aclarar las estrategias, organizar equipos y alinear las funciones (DAMA International, 2017, p. 37).

Teniendo en cuenta que la gestión de la información y de los datos es una ciencia en desarrollo, que además implica la adaptación continua y a largo plazo de toda la organización en sus distintas dimensiones, para que el Modelo de Gestión de la Información cumpla con su propósito, debe surtir múltiples adaptaciones a medida que se va desplegando su implementación, y los actores de la Universidad lo van interiorizando y apropiando. Es un fenómeno normal que toda institución humana orientada a la modificación del comportamiento de los actores sociales requiera modificaciones y perfeccionamiento de tal manera que su madurez se traduzca en un incremento de efectividad.

En este sentido, el Comité de Gobierno de la Información, o la dependencia encargada provisionalmente de ejercer el rol de gobierno de la información en la Universidad, deberá realizar ajustes al Modelo de Gestión de la Información, teniendo presente distinto tipo de factores. Algunos de estos factores pueden ser los siguientes:

- Preferencias y priorizaciones que se establezcan en el marco del Comité Nacional de Estrategia Digital.
- Directrices y lineamientos brindados desde la Rectoría a la Oficina de Gestión de la información de la DNED.
- Ejercicios de planeación que se realicen al interior de la Universidad.
- Pilotos que se realicen en el marco de la implementación del Modelo de Gestión de la Información.
- Acuerdos que se realicen en el marco del Clúster de Gestión de Información.

2.3.3.5. Formulación de Estrategia de Gobierno de la Información

El Comité de Gobierno de la Información o quien ejerza provisionalmente el rol de gobierno de la información antes de su constitución, deberá elaborar una estrategia de gobierno de datos e información. Esta estrategia deberá incluir los siguientes elementos:

- La carta o acta de constitución (visión, misión, principios, diagnóstico).
- Direccionar los actuales y futuros objetivos estratégicos organizacionales que se impulsarán a través de planes para el uso de los datos.

- Lineamientos u objetivos para aprovechar desarrollos tecnológicos y expectativas del mercado.
- Elaborar un esquema de rendición de cuentas de los datos.
- Dar cuenta de los aspectos específicos del gobierno de los datos (valor, riesgo, restricciones).
- Establecer un conjunto de expectativas (metas) que pueden requerir una revisión de la estrategia general para dar cuenta de nuevas oportunidades o riesgos.
- El marco operativo y las responsabilidades (estructura y deberes).
- La hoja de ruta de implementación (plazos, glosario, valoración de activos, claridades de los procesos, entregables o evidencias de apoyo).
- El plan de éxito operativo (expectativas sobre las actividades de gobierno).

2.3.4. Salidas

- Acta de conformación del Comité de Gobierno de GI
- Agenda de trabajo de Comité de Gobierno de GI
- Lista de temas priorizados
- Modelo de GI Ajustado y revisado
- Plan de trabajo GI

2.3.5. Usuarios

- Todos los actores del sistema de GI
- Comité de Gobierno de GI
- DNED – Oficina de GI
- Rectoría
- CNED
- Todos los actores del sistema GI

3. ARQUITECTURA DE DATOS

3.1. Conceptualización

3.1.1. Arquitectura de datos: definición

De acuerdo con DAMA, la arquitectura de datos es la organización fundamental de un sistema a nivel de los datos, incorporada en sus componentes, sus relaciones entre sí y con el ambiente, así como los principios que rigen su diseño y evolución. De acuerdo con lo anterior, la arquitectura de datos busca identificar las necesidades de datos de la empresa y diseñar y mantener los planos maestros para conocerlas. Desde la arquitectura de datos es fundamental:

- Definir el estado actual de los datos en la organización.
- Proporcionar un vocabulario organizacional estándar para datos y componentes.
- Alinear la arquitectura de datos con la estrategia empresarial y la arquitectura de negocio.
- Expresar los requerimientos estratégicos de datos.
- Delinear diseños integrados de alto nivel para cumplir con los requerimientos estratégicos de datos.

La arquitectura de datos se enmarca en un concepto más general denominado arquitectura empresarial, la cual tiene una visión holística de toda la organización y se lleva a cabo en 4 niveles diferentes:

- Al identificar cómo la organización crea valor para sus clientes y otros interesados, lo cual corresponde a la arquitectura de negocio empresarial, donde se definen los procesos misionales, entre otros elementos.
- Al describir cómo los datos deben estar organizados y administrados, lo cual corresponde a la arquitectura de datos.
- Al describir la estructura y funcionalidad de las aplicaciones en la empresa, lo cual corresponde a la arquitectura de aplicaciones empresarial.
- Al describir la tecnología física necesaria para habilitar el funcionamiento de los sistemas y entregar valor, lo cual corresponde a la arquitectura tecnológica empresarial.

3.1.2. Arquitectura de datos: elementos

En el marco metodológico del DAMA se identifica la cadena de valor de esta dimensión, tanto a nivel de artefactos de entrada, actividades y salidas, así como a nivel de actores y partes interesadas como los proveedores, participantes y consumidores de dichos artefactos. El diseño de arquitectura de datos, los diagramas de flujos de datos y la matriz CRUD (creación, lectura, actualización y borrado de datos) son

entregables fundamentales de la arquitectura de datos y se abordan en esta dimensión. Como se explicó en la Introducción del modelo, esta dimensión se estructuró a partir de la línea metodológica de DAMA (2017) y su aplicación será el contexto particular de la Universidad Nacional de Colombia.

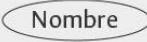
3.1.3. Arquitectura de datos: notación

Teniendo en cuenta que uno de los objetivos de la arquitectura de datos es la generación de planos maestros, dichos planos requieren una notación definida. A continuación, presentamos algunos temas centrales sobre los tipos de notación.

3.1.3.1. Notación entidad-relación tradicional

Definido originalmente por Peter Chen (1976), este es un modelo orientado al diseño de bases de datos a partir de una percepción del mundo real con una colección de objetos básicos como entidades y una diversidad de relaciones que se dan entre ellos (**Tabla 7**). Las entidades corresponden usualmente a sustantivos mientras que las relaciones corresponden a verbos. Desde su creación original se han desarrollado diversas extensiones al modelo, generando muchos otros descriptores de notación, los cuales se presentan con mayor detalle en la dimensión de Modelado y Diseño De Datos.

Tabla 7 Notación básica utilizada en el Diagrama Entidad-Relación

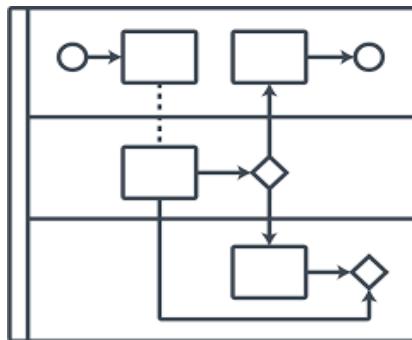
Descripción	Símbolo	Ejemplo
Rectángulos: representan conjuntos de entidades.	Entidad	
Elipses: representan atributos.	Atributo	
Líneas: conectan los atributos a los conjuntos de entidades, y los conjuntos de relaciones.	Conexión	
Rombos: representan relaciones.	Relación	

3.1.3.2. Notación diagrama de carriles (swimlane)

El diagrama de carriles (o *swimlane*) es un modelo identificado por Geary Rummler y Alan Brache (2010) en los años 90 para graficar y delimitar la pertenencia de procesos u otros objetos a unidades, organizaciones o áreas temáticas que corresponden a las líneas del diagrama (**Figura 5**), de manera

análoga a una piscina olímpica donde cada nadador pertenece a uno de los carriles delimitados. Dicho diagrama es útil para la identificación de responsabilidades sobre los procesos u objetos correspondientes.

Figura 5 Diagrama de Carriles (Swimlane)

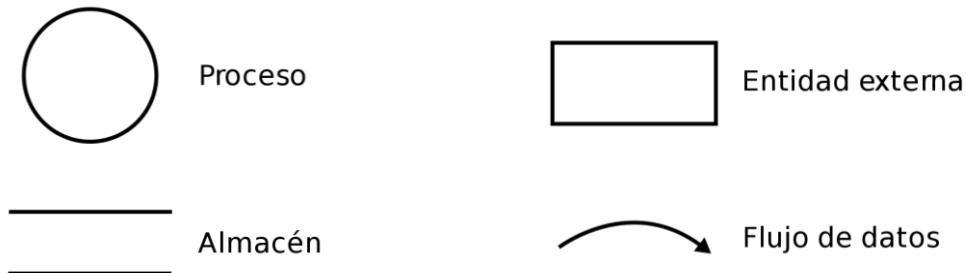


En esta dimensión se define un híbrido entre el modelo entidad-relación convencional y el diagrama de carriles como la aproximación más adecuada para representar la arquitectura de datos al interior de la Universidad. Lo anterior debido a que se identifican tanto las entidades como las relaciones principales, así como también su pertenencia a las diferentes áreas temáticas de la Universidad, correspondientes a los macroprocesos, sobre lo cual profundizaremos más adelante.

3.1.3.3. Notación Yourdon-De Marco para diagramas de flujos de datos

Otro de los artefactos importantes de esta dimensión es el diagrama de flujo de datos². Éste es una representación gráfica del flujo o movimiento de los datos desde un origen hacia un destino definido, pasando por el proceso al cual se realiza el análisis (**Figura 6**).

Figura 6 Componentes del diagrama de flujo de datos, notación Yourdon-De Marco



² Desarrollado originalmente por Larry Constantine con diversas ampliaciones en la notación, donde se destaca la Yourdon-DeMarco (1979).

Los flujos de datos son de naturaleza recurrente, siendo representados en diferentes niveles de profundidad. Se inicia con el nivel 0 (diagrama de contexto) el cual se va desagregando paulatinamente hasta niveles mayores de detalle de acuerdo con los lineamientos definidos en la organización que se presentan más adelante.

3.2. Propósito

En el marco del proyecto para la Transformación Digital es fundamental para la Universidad Nacional de Colombia plantear un modelo de arquitectura de datos para gestionar su información. En la Universidad, como es usual en las organizaciones de gran tamaño, se procesan más datos de los que las personas pueden comprender, por lo que es necesario representarlos en diferentes niveles de abstracción para que puedan entenderse y la alta dirección pueda tomar decisiones al respecto.

Históricamente en la Universidad se han realizado desarrollos enfocados en la solución de problemas específicos de las diferentes áreas, de donde han surgido como respuesta los sistemas de la información vigentes en la actualidad. Estos cumplen correctamente su función específica, pero, dado el crecimiento y la digitalización a nivel global, se hace difícil la generación de la información unificada con una misma semántica para todas las áreas de la Universidad. Esta dificultad en la unicidad de la información a reportar se ha visto reflejada en algunos problemas identificados en el Diagnóstico de Política de Gestión de la Información. Por lo tanto, esta dimensión tiene como propósito establecer el conjunto de componentes base de información, las relaciones entre ellos y su ubicación en el contexto de datos de la Universidad, así como elaborar otros artefactos que sirvan de guía para la estandarización de la información definida en el alcance para la fase I del modelo teniendo como horizonte el PLEI 2034.

3.2.1. Metas

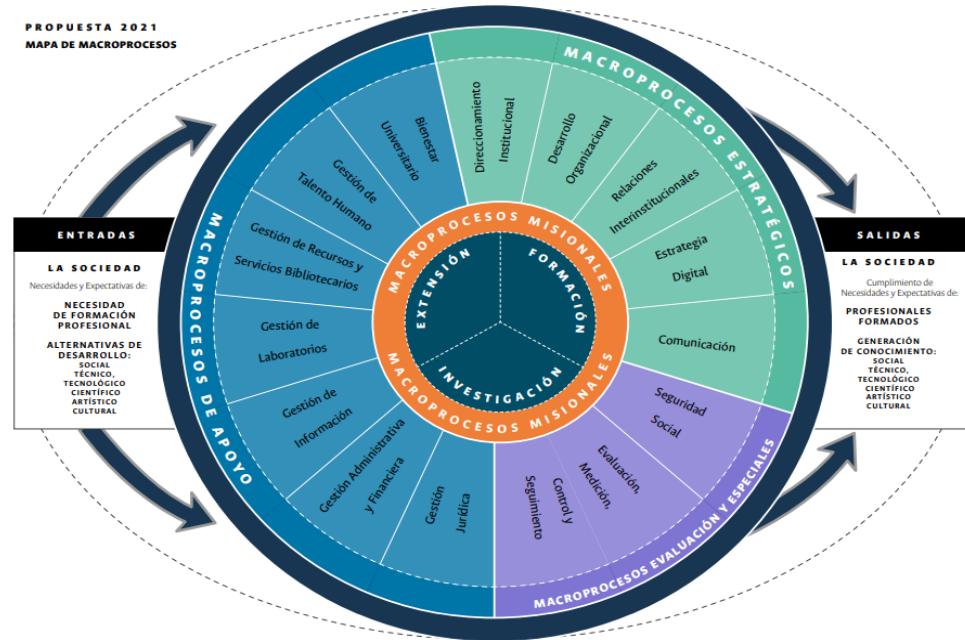
- Articular la arquitectura empresarial con la propuesta de la política de gestión de la información.
- Establecer las consideraciones metodológicas para la construcción de la arquitectura de datos de la universidad.
- Fortalecer los diseños de arquitectura de datos para la Universidad
- Definir un esquema de identificación regular de necesidades de la información de la Universidad dentro del alcance definido y las consideraciones metodológicas correspondientes.

3.2.2. Alcance fase II

El punto de partida al interior de la Universidad es el mapa de procesos vigentes que se presenta en la **Figura 7**. En él se pueden identificar los Macroprocesos como elementos fundamentales. Teniendo en cuenta su relevancia y su relación directa con las entidades de datos maestros, se adoptan para esta segunda fase los macroprocesos misionales de formación (extensión e investigación) y los macroprocesos de apoyo (gestión de bienestar y gestión del talento humano). La naturaleza del macroproceso de apoyo - gestión administrativa y financiera difiere de los demás macroprocesos del presente análisis, por esto, solamente se presentarán algunos elementos generales como partes

interesadas relacionadas, pero este macroproceso no se abordará en detalle dentro de esta segunda fase.

Figura 7 Mapa de Procesos de la Universidad Nacional de Colombia



3.3. Desarrollo de la fase II de arquitectura de datos

Tabla 8 SIPOC para arquitectura de datos aplicado a la Universidad

Proveedores	Entradas	Actividades	Salidas	Usuarios
Analistas de datos, dependencias priorizadas, administradores y custodios de datos	Instrumentos de recolección de datos sobre necesidades de información	Identificar las necesidades de información	Actualización de identificación de necesidades de información	DNED, Comité de Gobierno de Información
Arquitectos empresariales	Documentos de arquitectura empresarial	Articular con la arquitectura empresarial	Modelo de datos empresarial	Administradores de bases de datos, desarrolladores de software
DNED	Actualización de identificación de necesidades de la información	Construir el diseño de arquitectura de datos	Diseño de arquitectura de datos	DNED, Comité de Gobierno de la Información
DNED	Actualización de necesidades de información	Construir los diagramas de flujo de datos	Modelo de flujo de datos actualizado	DNED, Comité de Gobierno de la Información, procesos incluidos en modelo de flujo de datos

DNED	Actualización de necesidades de información, Diseño de arquitectura de datos	Construir la matriz CRUD	Matriz CRUD actualizada	Desarrolladores de software
------	--	--------------------------	-------------------------	-----------------------------

3.3.1. Proveedores

En este momento en la Universidad Nacional de Colombia no se encuentran definidos los roles de arquitectos empresariales; lo que se tiene son dependencias encargadas de estos contenidos. Respecto a las demás partes interesadas aquí mencionadas, sí existen en la Universidad.

- **Arquitectos de datos:** crean y mantienen el conocimiento organizacional sobre los datos y los sistemas a través de los cuales se mueven. Este conocimiento permite a una organización gestionar sus datos como un activo y aumentar el valor que obtiene de sus datos mediante la identificación de oportunidades para el uso de los datos, la reducción de costos y la mitigación de riesgos.
- **Arquitectos empresariales:** equipo encargado de los ejercicios de arquitectura empresarial. Las dependencias donde se deben encontrar los roles de arquitectos empresariales en la Universidad son: Vicerrectoría General - Sistema de Gestión Académico, Administrativa y Ambiental (SIGA) y Dirección Nacional de Estrategia Digital (DNED).
 - La Vicerrectoría General debe dar cumplimiento a dos objetivos específicos: garantizar la calidad, cualificación y transparencia institucional y liderar el desarrollo organizacional de la Universidad Nacional de Colombia.
 - La DNED debe cumplir los siguientes objetivos: diseñar y formular las políticas y la planeación estratégica de tecnologías de la información y las comunicaciones que apoyen los objetivos de la Universidad Nacional de Colombia para que sean adoptadas por el Comité Nacional de Tecnologías de la Información y las Comunicaciones (CNTIC) y definir los estándares y lineamientos sobre Tecnologías de la Información y las Comunicaciones (TIC) para toda la Universidad Nacional de Colombia.
- **Administradores y custodios de datos:** administradores de los sistemas de la información en la Universidad. Los sistemas de la información que se tendrán en cuenta en esta dimensión son: Sistema de la información Académica (SIA), Sistema de la información de la Investigación, Extensión y Laboratorios (HERMES), Sistema de la información de Bienestar Universitario (SIBU), Sistema de la información de Talento Humano (SARA) y Sistema de la información administrativo y financiero (QUIPU).
- **Expertos de los temas de negocio y analistas de datos:** líderes funcionales y asesores de las áreas misionales. En el macroproceso de Investigación y Extensión corresponde a la Vicerrectoría de Investigación y Extensión. Para el macroproceso de Formación la dependencia es Vicerrectoría Académica y para los macroprocesos de Bienestar y Talento Humano la dependencia es la Vicerrectoría General. Para el macroproceso de Bienestar, la Dirección Nacional de Bienestar

Universitario y finalmente, para el macroproceso de Gerencia Financiera y Administrativa la dependencia es Rectoría junto con la Gerencia Financiera y Administrativa.

3.3.2. Entradas

- Documentos de arquitectura empresarial mencionados en los antecedentes.
- Arquitectura de negocios: se tienen en cuenta tanto el mapa de procesos de la Universidad, así como los documentos de caracterización de dichos procesos.
- Levantamiento de necesidades de información.

3.3.3. Actividades

3.3.3.1. Identificar las necesidades de información

Se debe analizar con los expertos de cada área o macroprocesos qué información o datos requieren tener para suplir las necesidades de la información encontradas. Esto con el objetivo de conocer quién es el custodio de la información y con cuántas áreas se comparte la misma necesidad de información.

Para el conjunto de necesidades de la información se han tenido en cuenta la metodología de IT4+ (MinTIC, 2019a) y la guía sobre cómo construir el catálogo de componentes de la información (MinTIC, 2016). Inicialmente se tomaron como base los nomogramas de los procesos a revisar y se identificaron los potenciales atributos que se requerían. Dado que este proceso puede ser bastante extenso en los análisis con restricciones importantes de tiempo (aunque puede aportar más información sobre el diagnóstico de necesidades), se tuvo como alternativa abordar las necesidades con entrevistas directas a los expertos de cada área, agrupándolas en categorías a nivel de proceso. La metodología para el levantamiento de necesidades se presenta en detalle en el **Anexo 1 - Cruce de la información disponible y necesidades existentes**. Para realizarla se tuvieron en cuenta los siguientes pasos:

1. Establecer los macroprocesos donde se debe identificar la información y las necesidades de acuerdo con la priorización definida por el clúster de gestión de la información y su nivel de granularidad, que en el caso de fase I, se realizó a nivel de proceso. Para esta primera fase se definieron los relacionados en la **Tabla 9**.

Tabla 9 Macroprocesos y procesos identificados para el análisis de necesidades - Fase I

Macroproceso	Proceso	Sistema de la información relacionado
M1 Misionales – Formación	M1PR1 Gestión Administrativa de Apoyo a la Formación- GAAF. <i>Etapa Gestionar Admisiones, registro y matrícula en el proceso de GAAF</i>	Sistema de la información de Admisiones

Macroproceso	Proceso	Sistema de la información relacionado
M1 Misionales – Formación	M1 PR1 Gestión Administrativa de Apoyo a la Formación- GAAF. <i>Etapa Gestionar programas de pregrado y postgrado en el proceso de GAAF</i>	Sistema de la información Académica - SIA
M1 Misionales – Formación	M1 PR1 Gestión Administrativa de Apoyo a la Formación- GAAF. <i>Etapa de Gestionar admisiones, Registro y matrícula en el proceso de GAAF</i>	Sistema de la información Académica - SIA
M2 Misionales - Investigación y Creación Artística	M2PR1 Investigación y Creación Artística	Sistema de la información HERMES - Investigación
M3 Misionales - Extensión	M3PR1 Gestión de la Extensión	Sistema de la información HERMES - Extensión
M4 Apoyo - Bienestar Universitario	M4PR1 Bienestar Universitario	Sistema de la información de Bienestar Universitario - SIBU
M9 Apoyo - Gestión de Talento Humano	M9PR1 Gestión de Talento Humano	Sistema de la información de Talento Humano - SARA

2. Realizar sesiones de trabajo con los profesionales encargados de la parte funcional, de los flujos de la información y de los sistemas de la información relacionados con los macroprocesos/procesos del punto anterior.
3. Establecer sesiones de trabajo con los profesionales mencionados en el numeral anterior, encargados de dar respuesta a los requerimientos del Ministerio de Educación Nacional para los datos que se entregan al Sistema Nacional de la información de la Educación Superior (SNIES).
4. En cada sesión de trabajo anteriormente descrita se identificarán:
 - o Flujos de la información existentes, entrantes y salientes de cada macroproceso. Se buscaron requerimientos que se suplen actualmente entre dependencias internas y/o entidades externas a la Universidad como respuesta a necesidades que han evidenciado y que cada proceso puede suplir con los datos que gestiona.

- Necesidades de la información que, a la fecha, no se han podido suprir o se suplen parcialmente, identificando demanda no satisfecha y las razones que han llevado a que no pueda cumplirse con este requerimiento.
- Los grupos de datos que confirman la información y las necesidades derivadas, los macroprocesos/procesos que producen y consumen la información, su frecuencia, medio y elementos de mejora de cada uno.

3.3.3.2. Articular con Arquitectura Empresarial

Se puede hacer un paralelo entre lo realizado en la metodología TOGAF de arquitectura empresarial y el modelo DAMA señalando que la fase B (arquitectura de negocios) de TOGAF corresponde al dominio de arquitectura de negocio empresarial del DAMA. Por su parte, la fase C, arquitectura de sistemas de la información TOGAF, corresponde en esta dimensión al dominio de arquitectura de aplicaciones empresarial DAMA en donde también se hace referencia a elementos del dominio de arquitectura de datos empresarial. Finalmente, la fase D, arquitectura de tecnología TOGAF, corresponde al dominio de arquitectura tecnológica empresarial DAMA. Estos paralelos se presentan en la **Tabla 10**.

Tabla 10 Paralelo entre la metodología TOGAF-ADM y la metodología DAMA y su articulación con la arquitectura empresarial

Metodología TOGAF - ADM de Arquitectura Empresarial	Metodología DAMA, dimensión de Arquitectura de Datos
Fase B. Arquitectura de Negocios	Dominio de arquitectura de negocio empresarial
Fase C. Arquitectura de Sistemas de Información	Dominio de arquitectura de aplicaciones empresarial
	Dominio de arquitectura de datos empresarial
Fase D. Arquitectura de Tecnología	Dominio de arquitectura tecnológica empresarial

La Metodología TOGAF propone algunos artefactos asociados para la fase de sistemas de la información en la arquitectura de datos como el catálogo de entidades de datos/componentes de datos, la matriz entidades de datos/funciones de negocio, el diagrama conceptual de datos, el diagrama lógico de datos, el diagrama de diseminación de datos, entre otros, los cuales se abordan en algunas dimensiones de la metodología DAMA.

De manera general, cuando se realiza la construcción de una arquitectura de datos, es necesario identificar la conexión de la metodología correspondiente con la arquitectura empresarial previamente existente. De no existir, se debe mencionar la pertinencia de crear las otras 3 áreas (arquitectura de negocio -u organizacional-, arquitectura de sistemas de la información y arquitectura de tecnologías de información).

- **Identificación de necesidades de información:** son canalizadas a través de los servicios de información. En el modelo se identifican con el prefijo 'N' seguido del consecutivo y el nombre; por ejemplo: *N1 Es necesario contar con los campos de estrato y núcleo familiar de todos los estudiantes de pregrado y posgrado.*
- **Servicios de información:** corresponde a la manera en que se canalizan y presentan las necesidades de información. En el modelo se identifican con el prefijo 'SE' seguido del consecutivo y el nombre; por ejemplo: *SE1 Estadísticas.*
- **Identificación de la información** que interopera entre las áreas de negocio. En el modelo se identifican con el prefijo 'IN', seguidos del consecutivo y el nombre correspondiente; por ejemplo: *IN1 Porcentaje de incremento salarial para la siguiente vigencia.*

3.3.3.3. Construir el diseño de arquitectura de datos

Una vez identificados los componentes base de la información es necesario articularlos entre sí teniendo en cuenta ciertos criterios, a saber:

- Las entidades maestras son de relevancia transversal a las diferentes áreas temáticas.
- Las tablas de referencia son utilizables por cualquiera de los demás componentes de información.
- Tanto los datos intrínsecos de las entidades de datos, así como los datos transaccionales producto de las relaciones, están encaminados a satisfacer alguna de las necesidades de información.
- Los servicios de la información corresponden a la manera de canalizar la información que se identifica en las necesidades de información.

El diseño de arquitectura de datos es el diseño abstracto que conecta los diferentes componentes base de información. Teniendo en cuenta que es un plano maestro, este diseño debe construirse como un diagrama fácilmente legible para diversas audiencias. En el caso de la Universidad corresponde a una combinación de un modelo entidad-relación tradicional con un diagrama de funciones cruzadas. Su objeto es visualizar las relaciones que se dan entre las entidades y delimitar claramente las áreas temáticas que, en el caso de la Universidad, corresponden a los macroprocesos.

En el costado izquierdo del diseño se ubican las entidades maestras con los atributos de la información más relevantes y transversales a dos o más áreas temáticas/macroporcesos, junto con las tablas de referencia y otras entidades no maestras que aporten información estructurada. De igual manera, existe un conjunto de la información no estructurada (por ejemplo, redes sociales y lagos de datos), que complementa la información procesada por la Universidad.

En el centro del diagrama se ubican las áreas temáticas de la Universidad que corresponden a los macroporcesos misionales de formación, investigación y extensión, así como también a dos macroporcesos de apoyo (bienestar y gestión humana), los cuales hacen parte de la fase I del proyecto. Adicionalmente se menciona aquí el macroporceso de gestión financiera por su importancia para la Universidad y debido a que contiene algunas de las necesidades de la información identificadas (su detalle se abordará en fases posteriores del proyecto).

Bajo las áreas temáticas se encuentran las relaciones entre las entidades maestras de la izquierda (filas) y las áreas temáticas de la Universidad (columnas) correspondientes. Dichas relaciones alojan información transaccional de operaciones, teniendo un identificador compuesto proveniente tanto de la entidad maestra como del área temática correspondiente. Continuando hacia la derecha, la información generada por estas relaciones pasa a suplir un conjunto de necesidades de información, las cuales se canalizan a través de los servicios de información. La propuesta del diseño de arquitectura de datos se presenta en la sección 3.3.5.1 - Diseño de arquitectura de datos.

3.3.3.4. Construir los diagramas de flujos de datos

Los flujos de datos se presentan en varios diagramas con diferentes niveles de profundidad. En ellos se inicia con el nivel 0 (diagrama de contexto), identificando a la Universidad completa como un gran Macroproceso que recibe y genera datos e información. El nivel 1 de los flujos de datos corresponde a los Macroprocesos definidos en el alcance (formación, extensión, investigación, bienestar, gestión de talento humano) y el nivel 2 corresponde a los procesos, donde se seleccionaron los más relevantes para esta fase.

Para la construcción de los diagramas de flujos de datos se usó como insumo de trabajo la matriz de necesidades de la información desarrollada en la etapa inicial del levantamiento de información. A partir de ella se identificaron las necesidades asociadas a cada macroproceso y atributos como el tipo de la información que contiene la necesidad, el área productora y consumidora de dicha información, la frecuencia con la que se presenta la necesidad, entre otros. Lo anterior con el objetivo de identificar qué tipo de la información se entrega dentro de cada macroproceso y qué se obtiene como salida o resultado de la información procesada.

3.3.3.5. Construir la matriz CRUD (entidades/procesos)

Este es un método para identificar los lugares donde las entidades de datos más relevantes tienen participación, es decir, dónde se da su creación, lectura, actualización y borrado. Este artefacto permite ver con mayor claridad qué proceso es dueño de la entidad de datos respectiva y tiene el gobierno sobre ella. En un eje se listan las entidades de datos correspondiente y en el otro eje se listan las áreas temáticas que se hayan elegido.

Para el caso concreto de la Universidad, en las filas se enlistaron las entidades maestras (entidades de datos con mayor relevancia) y en las columnas se ubicaron los procesos más relevantes de acuerdo con el alcance definido. Las entidades maestras de datos se ubicaron en las filas y los procesos del alcance se ubicaron en las columnas respectivas. Las letras permiten identificar en dónde se crean, leen, actualizan y borran los datos de cada entidad. El resultado de la presente actividad se presenta en la sección 3.3.5.5 - Matriz CRUD.

3.3.4. Participantes

- **Arquitectos de Datos Empresariales:** futuro grupo de arquitectura de datos que estará encargado de las siguientes actividades:

- Coordinar el diseño de procesos y procedimientos que permitan gestionar apropiadamente las necesidades del ciclo de vida completo de los datos de la Universidad.
 - Proyectar conceptos sobre la arquitectura de datos en la Universidad, apoyando la toma de decisiones sobre proyectos que puedan transformar políticas, definiciones o los datos mismos.
 - Estudiar requerimientos de la información por parte de las áreas de la Universidad para garantizar que estén de acuerdo con las políticas y estándares definidos para la gestión de datos.
 - Apoyar la socialización de las políticas y lineamientos definidos en el marco de la arquitectura de datos para que sean cumplidas por parte de todas las áreas de la Universidad.
 - Realizar el seguimiento a las nuevas tendencias y buenas prácticas que apliquen a la arquitectura empresarial dentro del área de su competencia.
 - Relacionado con los temas de arquitectura, se encuentra a nivel técnico, el rol de Administrador de Base de Datos (DBA).
- **Modeladores de Datos:** futuro grupo de modelado de datos. Los roles relacionados serán:
- **Diseñador de base de datos:** modelado de datos, diseño de base de datos, técnicas de análisis y diseño orientado a objetos.
 - **Diseñador de interfaz:** este rol no es responsable de implementar la interfaz de usuario, se centra en el diseño y la "forma visual" de la interfaz de usuario capturando requisitos, construyendo prototipos de dicha interfaz, incluye a otros interesados en revisiones de utilización y sesiones de prueba de uso, revisando y proporcionando la información de retorno apropiada en la implementación final.
 - **Analista de sistemas:** se espera que este rol tenga experticia en la identificación y la comprensión de problemas y oportunidades; esto incluye la capacidad de articular las necesidades que se asocian con el problema clave que se debe solucionar o la oportunidad a realizar. El conocimiento de la Universidad y dominio de la tecnología son habilidades adicionales útiles para quienes desempeñan este rol.

3.3.5. Salidas

Producto de las actividades realizadas se generaron las siguientes salidas:

3.3.5.1. Diseño de arquitectura de datos

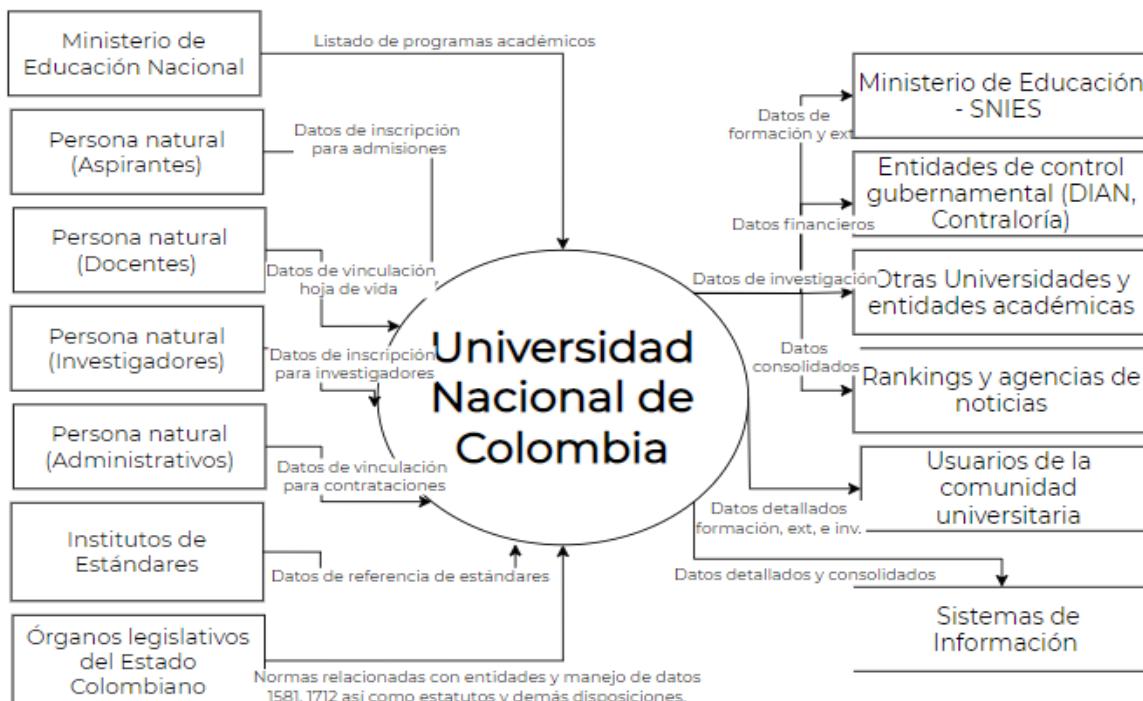
Es el diseño abstracto que conecta los diferentes componentes base de información. En el caso de la Universidad corresponde a una combinación de un modelo entidad-relación tradicional, con un diagrama de funciones cruzadas. Su objeto es visualizar las relaciones que se dan entre las entidades y delimitar claramente las áreas temáticas que, en el caso de la Universidad, corresponden a los macroprocesos.

3.3.5.2. Diagramas de flujos de datos

De acuerdo con las consideraciones metodológicas mencionadas en la actividad 3.3.3.4, la **Figura 8** se presenta una propuesta del diagrama de contexto, identificando a la Universidad como un gran Macroproceso. En este diagrama los rectángulos representan entidades externas/organizaciones/dependencias; es decir, sujetos que envían o reciben información. Las flechas corresponden a los objetos de la información que es transmitida; dicha información es “cruda” en las flechas de la izquierda y “procesada” en las flechas de la derecha. El óvalo del proceso debe entenderse como una “cocina” que recibe datos crudos y genera datos procesados.

Posteriormente se realizó un primer nivel de desagregación teniendo en cuenta el alcance de la primera fase. En él se incluyeron los macroprocesos misionales de la Universidad: formación, extensión e investigación y los macroprocesos de apoyo (bienestar y gestión del talento humano). Dichos flujos de datos se ubicaron en el Diagrama de flujo de datos nivel 1³. En algunos casos específicos se llegó al detalle a nivel de proceso que corresponde al nivel 2 de flujos de datos.

Figura 8 Diagrama de flujo de datos nivel 0



³ Disponible en: t.ly/Zqiw

3.3.5.3. Modelo de datos empresarial

El modelo empresarial de datos, en este caso, corresponde a una implementación del diseño de arquitectura de datos. Para esto, se puede remitir al **Anexo 2 – Diagrama de entidades maestras**, que contiene las entidades maestras identificadas en la primera fase. En este punto, esta dimensión enlaza con la dimensión de Modelado y Diseño de datos, donde se realizó una desagregación del modelo de datos empresarial en diagramas conceptual y lógico (Ver sección 4.3.3.3.1 del capítulo MODELADO Y DISEÑO DE DATOS).

3.3.5.4. Levantamiento de necesidades de información

Para este procedimiento se construyó el **Anexo 3 - Matriz Necesidades de la información** y se realizó el cruce con la información existente, identificando la demanda satisfecha e insatisfecha. En la **Tabla 11** se presentan dos ejemplos de las necesidades identificadas.

Tabla 11 Ejemplo de necesidades de la información por categorías a nivel de proceso (Fase 1 del proyecto)

Código identificador de la necesidad	Necesidad de la información (Categorizadas a nivel de proceso en la fase 1 del proyecto)	Tipo de Macroproceso - Nombre Macroproceso / Proceso/ Etapa
N1	Es necesario contar con los campos de estrato, núcleo familiar, discapacidades y etnias de todos los estudiantes de pregrado y posgrado	Misionales- Formación / Gestión administrativa de apoyo a la formación/etapa Registro y Matrícula
N2	Es necesario contar con la identificación de los estudiantes, nombres completos, direcciones, teléfonos y correos electrónicos	Misionales- Formación / Gestión administrativa de apoyo a la formación/etapa Registro y Matrícula

3.3.5.5. Matriz CRUD

Teniendo en cuenta las consideraciones de la actividad 3.3.3.5, se construyó la matriz CRUD presentada en el **ANEXO 4_Matriz CRUD Identidades-Procesos**. En la **Figura 9** presentamos, a manera de ejemplo, una sección de dicha matriz. En ella se puede ver que las entidades maestras de datos se ubican en las filas y los procesos del alcance se ubican en las columnas respectivas. Las letras permiten identificar en dónde se crean, leen, actualizan y borran los datos de cada entidad.

Figura 9 Sección de la Matriz CRUD

	A	B	C	D	E	F
1			Misionales - Formación / Gestión Administrativa de apoyo a la Formación/ Admisiones	Misionales - Formación / Apoyo a la Gestión de la Actividad Académica /Apoyo a la Innovación Académica	Misionales - Formación / Apoyo a la Gestión de la Actividad Académica / Gestión de la actividad académica	Misionales - Formación / Apoyo a la Gestión de la Actividad Académica / Gestión de Programas Curriculares
2		Programa Curricular		R		R C U D
3		Plan de Estudios		R		R C U D
4		Asignatura		R		R C U D

3.3.6. Usuarios

- Administradores de Base de Datos:** Ingenieros DBA internos y externos de los proveedores. La Universidad cuenta con sistemas de la información oficial contratados con externos y otros desarrollados internamente. Existe en cada caso una dependencia responsable de dichos sistemas de la información que generan los datos que se requieren para la toma de decisiones, así:

SARA: Dirección Nacional de Personal Académico y Administrativo

SIA: Dirección Nacional de la información Académica

QUIPU: Gerencia Nacional Administrativa y Financiera

SIBU: Dirección Nacional de Bienestar

HERMES: Vicerrectoría de Investigación

Actividades para tener en cuenta:

- Gestión General de Base de Datos
- Modelado de Datos y Diseño de Base de Datos
- Auditoría
- Integración con aplicaciones
- Resguardo y recuperación de datos
- Inteligencia de negocios y almacenamiento de datos
- Planificación de capacidad - Administración de cambios
- Apoyo en el desarrollo de aplicaciones (relacionadas con el manejo de la base de datos)

- Desarrolladores de Software:** Clúster de Sistemas de la información en la Universidad. La Universidad contrata en su mayoría con proveedores externos y pocas veces se realizan desarrollos internos, por lo anterior conviene definir un equipo para realizar desarrollos personalizados para la UNAL externos o internos (creación de la casa de software) que realice las siguientes actividades:

- Crear y desarrollar nuevos programas o sistemas.
- Evaluar sistemas nuevos y existentes.
- Mejorar programas existentes.
- Realizar el mantenimiento correspondiente en los sistemas existentes y el monitoreo y corrección de los defectos detectados.
- Elaborar manuales operativos y especificaciones técnicas de los sistemas.
- Trabajar en conjunto con el resto del equipo, es decir, con gerentes de proyecto, diseñadores gráficos, otros desarrolladores, administradores de bases de datos.
- Elaborar informes sobre el progreso del proyecto.

3. Gerentes de proyectos: líderes de proyectos relacionados con gestión de información. Existen dos roles en esta parte de interesados: gerente de proyectos técnico y gerente de proyecto funcional. El primero debe estar en la casa de software de la UNAL y el segundo es el director o persona encargada del proyecto. Actividades para tener en cuenta:

- Definición y presentación del proyecto
- Planificación del proyecto
- Establecimiento de objetivos
- Supervisión de tareas
- Implementación de soluciones o cambios

4. Equipos de soporte: El equipo de soporte debe estar definido de acuerdo con los desarrollos implementados.

4. MODELADO Y DISEÑO DE DATOS

4.1. Conceptualización

El modelado de datos es el primer paso en el proceso de diseño de bases de datos (Watt, 2014). Puede abarcar, entre otros, fases de diseño abstracto, documentación de bases de datos existentes, definiciones de requerimientos para nuevas aplicaciones. El modelado y diseño de datos busca producir una representación visual llamada “Modelo de Datos” (DAMA Internacional, 2017). Los modelos de datos se utilizan para representar los datos, definir su almacenamiento y establecer sus relaciones. Los modelos pueden ser de diferentes tipos dependiendo de su nivel de abstracción y propósito y aportan a la mejora de la gestión de la información en:

- Ofrecer una visión holística de los datos que se tienen en la institución.
- Garantizar la coherencia de los datos en cuanto a las convenciones de nomenclatura, los valores predeterminados, la semántica, entre otros, al tiempo que contribuyen con su calidad.
- Enfatizar qué datos se necesitan, cómo deben organizarse y cómo deben implementarse.
- Diseñar bases de datos en niveles conceptual, lógico y físico.
- Proporcionar una imagen clara de los datos que sirva a la vez como insumo para el desarrollo de nuevas bases de datos.
- Identificar datos faltantes y datos redundantes.
- Ayudar a definir tablas relacionales, llaves primarias y foráneas, procedimientos y tablas de referencia.

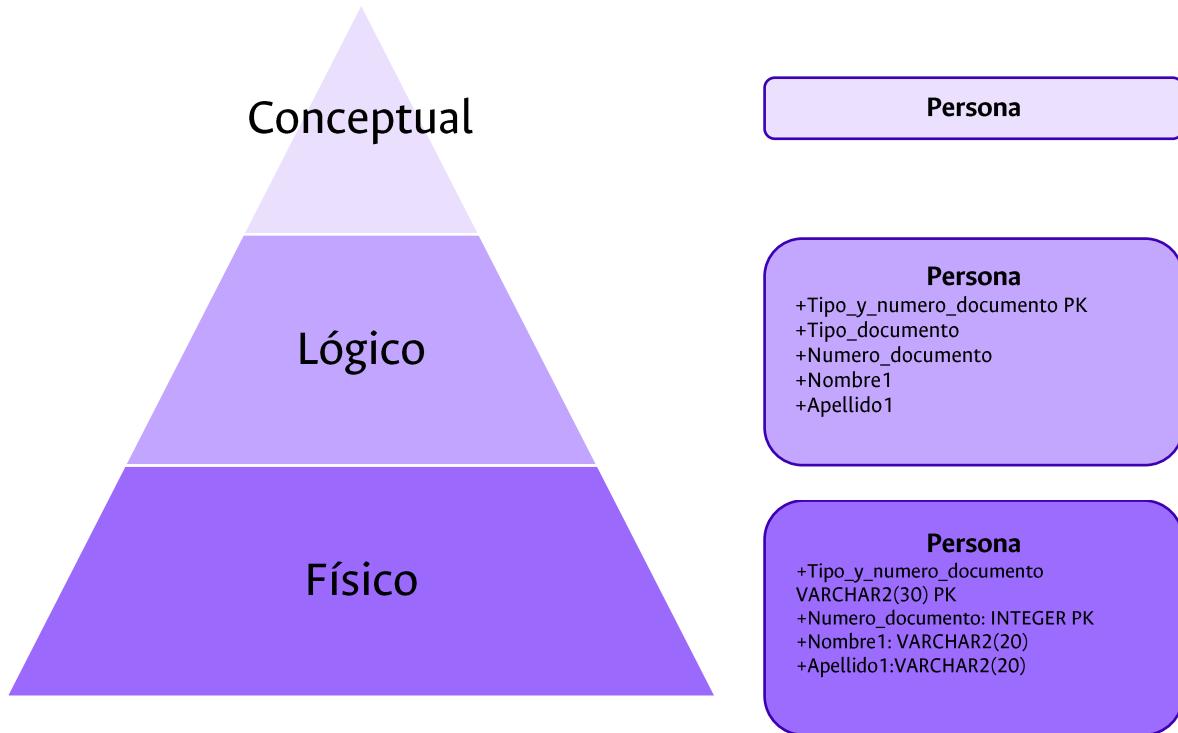
4.1.1. Tipos de modelos de datos

Existen principalmente tres tipos diferentes de modelos de datos: conceptuales, lógicos y físicos (**Figura 10**). Cada uno tiene un propósito específico:

- **Modelo de datos conceptual:** este modelo de datos define QUÉ contiene el sistema de la información (Taylor, 2021). El propósito es organizar, ampliar y definir conceptos y reglas aplicadas a los datos (Ansari, 2017).
- **Modelo de datos lógico:** define CÓMO se debe implementar el sistema independientemente del manejador de base de datos (Taylor, 2021). El propósito es desarrollar un mapa técnico de reglas y estructuras de datos que identifique los atributos y relaciones entre entidades (Ansari, 2017).
- **Modelo de datos físico:** este modelo de datos describe CÓMO se implementará el sistema utilizando un manejador de base de datos específico (Taylor, 2021). Define la estructura de

base de datos, así como detalles técnicos relacionados con el diseño de la aplicación (Ansari, 2017).

Figura 10 Tipos de modelos de datos y ejemplo, adaptación de Ansari, (2017)

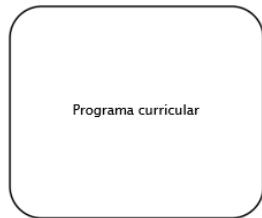


4.1.1.1. Representaciones usadas en modelos de datos conceptuales

Los modelos de datos conceptuales aportan una vista organizada y simple de los datos estableciendo entidades y sus relaciones; además, se generan de forma independiente a las especificaciones de hardware, software, almacenamiento de datos o manejador de base de datos (Taylor, 2021). Las diferentes representaciones para cada componente de un modelo conceptual se realizan de la siguiente manera:

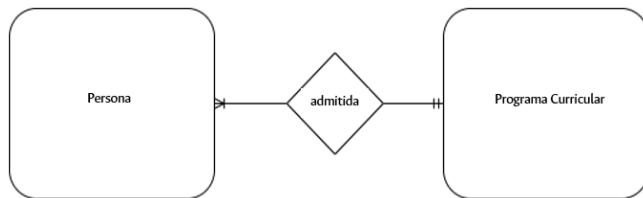
- a) **Entidades:** las entidades se representan con rectángulos de esquinas rectas o redondeadas y el respectivo sustantivo asociado. Por ejemplo, la entidad PROGRAMA CURRICULAR de la **Figura 11**.

Figura 11 Representación gráfica de entidades en el modelo conceptual



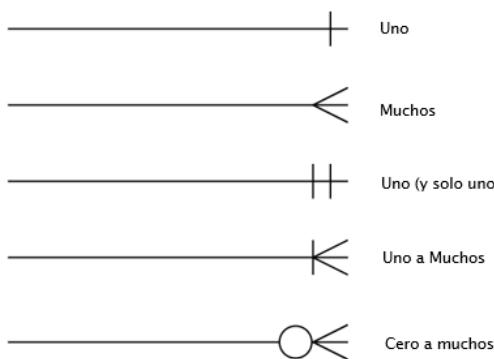
- b) **Relación:** las relaciones se representan con rombos en los que se escriben las asociaciones identificadas. Estas inician con un verbo en infinitivo que describe la transacción a realizarse entre la(s) entidad(es). Por ejemplo, *admitir* PERSONA en PROGRAMA CURRICULAR en la **Figura 12**.

Figura 12 Representación de relaciones en el modelo conceptual



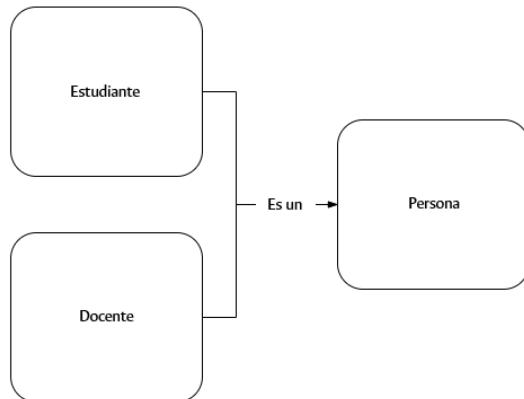
- c) **Cardinalidad y ordinalidad:** la cardinalidad se representa por un símbolo que unirá a las entidades o relaciones. Por ejemplo, una PERSONA es *admitida* a un PROGRAMA CURRICULAR y un PROGRAMA CURRICULAR puede *admitir* a muchas PERSONAS (**Figura 13**).

Figura 13 Representación de cardinalidad y ordinalidad del modelo conceptual



- d) **Herencia:** la herencia se representa con la frase “*es un*” sobre la línea que une la entidad padre e hija. Por ejemplo, un docente es una PERSONA y un estudiante es una PERSONA (**Figura 14**).

Figura 14 Representación de herencia en el modelo conceptual



- e) **Agregación o composición:** se representa con la frase “es parte de” sobre la línea que une a las entidades y establece una relación de agregación o composición. Por ejemplo, PLAN DE ESTUDIO es parte de PROGRAMA CURRICULAR (**Figura 15**).

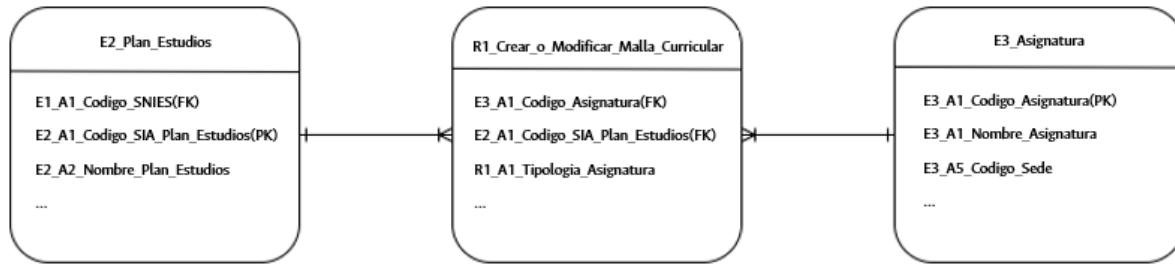
Figura 15 Representación de agregación o composición en el modelo conceptual



4.1.1.2. Representaciones usadas en modelos de datos lógicos

Las representaciones en el modelo de datos lógicos se utilizan para definir estructuras de los elementos de datos y establecer relaciones entre ellos. Además, agregan información adicional de la que aporta el modelo de datos conceptual y proporcionan una base para la generación del modelo de datos físico (Taylor, 2021). El modelo de datos lógico se construye de forma independiente del motor de base de datos; sin embargo, aporta con exactitud los atributos que intervienen en la definición de cada entidad o que se producen en las transacciones que se realizan. En este modelo pueden especificarse las llaves de cada entidad o que las intervienen en las relaciones. Por ejemplo, la transacción para crear o modificar una malla curricular en un PLAN DE ESTUDIOS de la **Figura 16**.

Figura 16 Representación de entidades y relaciones en el modelo lógico



La naturaleza de un dato se define en el modelo lógico mediante el uso de niveles de privacidad. Estos se representan con símbolos, como se muestra en la **Tabla 12**.

Tabla 12 Símbolos con su significado, utilizados en el modelo lógico

Símbolo	Significado
+	Representa datos públicos
#	Representa datos semiprivados
-	Representa datos privados
~	Representa datos sensibles

Como se indicó anteriormente, en el modelo lógico se especifican las llaves de cada entidad o aquellas que intervienen en las relaciones. Estas llaves se identifican como:

- **Llave primaria (PK):** es un atributo que identifica de forma única a una instancia de una entidad.
- **Llave foránea (FK):** es un atributo creado de manera implícita, cuando existe una relación de dependencia o agregación entre dos entidades y cuyo valor corresponde al de una llave primaria de alguna de las entidades involucradas en la relación.

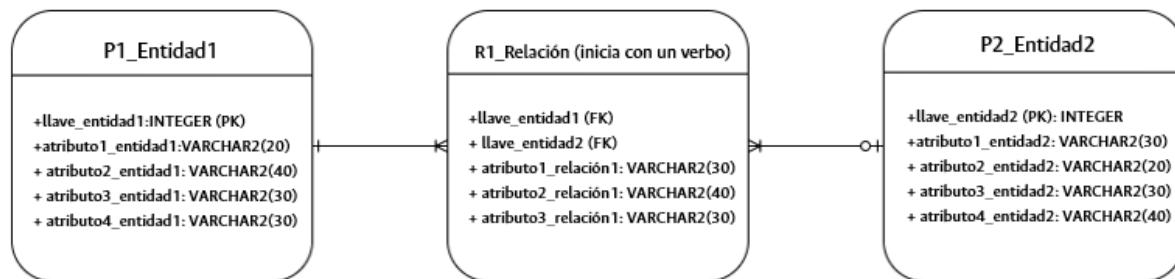
4.1.1.3. Modelo de datos físicos

El modelo de datos físicos describe la implementación física de los datos en la base de datos, aporta una visión abstracta de la(s) base(s) de datos en un solo esquema y contiene una especificación mayor que la que pueden aportar los modelos de datos conceptuales y lógicos (Taylor, 2021). Este modelo se desarrolla para una versión específica de manejador de base de datos, almacenamiento de base de datos y tecnología disponible. Además, representa como se construirá el modelo en la base de datos de acuerdo con sus restricciones y/o requisitos físicos. Este modelo se caracteriza ya que las entidades y relaciones identificadas previamente en los modelos conceptual y lógico son convertidas en tablas, los atributos se muestran en forma de columnas, especificando sus correspondientes tipos de datos y longitudes, como se representa en la **Tabla 13**, y, además, indica las llaves primarias y foráneas en las tablas de las relaciones, ver **Figura 17**.

Tabla 13 Ejemplos de tipos de datos presentes en el modelo físico

Tipo de dato	Descripción	Rango/Formato de almacenamiento	Longitud
Int, Integer	Número entero con o sin signo	Desde -2147483648 a 2147483647. Sin signo el rango va desde 0 a 429.4967.295	4 bytes
Float	Permite almacenar pequeños números decimales (de punto flotante)	Desde -3.402823466E+38 a -1.175494351E-38, 0 y desde 1.175494351E-38 a 3.402823466E+38	4 bytes
Double	Número en coma flotante de precisión doble	Desde -1.7976931348623157E+308 a -2.2250738585072014E-308, 0 y desde 2.2250738585072014E-308 a 1.7976931348623157E+308	8 bytes
Date	Tipo fecha, almacena una fecha	El rango de valores va desde el 1 de enero del 1001 al 31 de diciembre de 9999. El formato de almacenamiento es de año-mes-día	3 bytes
Datetime	Combinación de fecha y hora	El rango de valores va desde el 1 de enero del 1001 a las 0 horas, 0 minutos y 0 segundos al 31 de diciembre del 9999 a las 23 horas, 59 minutos y 59 segundos. El formato de almacenamiento es de año-mes-día horas:minutos:segundos	8 bytes
Char(n)	Almacena una cadena de longitud fija	La cadena podrá contener desde 0 a 255 caracteres	n bytes
Varchar(n)	Almacena una cadena de longitud variable	La cadena podrá contener desde 0 a 255 caracteres	n+1 bytes

Figura 17 Representación de entidades y relaciones en el modelo físico



4.2. Propósito

La dimensión de modelado y diseño de datos busca, primero, establecer un vocabulario común y una herramienta de comunicación alrededor de los datos; segundo, proveer una forma de documentación del conocimiento acerca de los datos, la información y los sistemas de la información de la Universidad. Tercero, proporcionar un instrumento para la integración de las aplicaciones existentes y futuras y, finalmente, recoger una visión holística de los datos de la institución aportando elementos que mejoran su comprensión y aprovechamiento (Ansari, 2017).

Como consecuencia del desarrollo de modelado y diseño de datos se identifican los principales elementos para la gestión de las dimensiones de datos maestros, datos de referencia, metadatos y gobierno de los datos. Además, se aporta a la comprensión de los datos existentes y a la forma en que se encuentran dentro de cada sistema de la información aumentando las oportunidades de integración e interoperabilidad entre sistemas de la información existentes y delimita el alcance para la creación de nuevos proyectos ya que identifica los datos existentes, su definición técnica y conceptual y su uso (DAMA Internacional, 2017).

4.2.1. Metas

- Promover la descripción centralizada y única los datos que se gestionan y almacenan en cada una de las fuentes de la información existentes de acuerdo con el alcance establecido.
- Identificar riesgos de redundancia de datos y duplicidad de esfuerzos desde varias dependencias.
- Documentar de forma explícita, y bajo una misma metodología, los datos gestionados por la Universidad, promoviendo el trabajo colaborativo entre las dependencias líderes de fuentes de información.

4.2.2. Alcance fase II

El desarrollo del modelado y diseño de los datos de los sistemas de la información corporativos de la Universidad es una actividad permanente y continua que se inició en la fase I del proyecto de gestión de la información y debe mantenerse y actualizarse periódicamente, de tal manera que se apropie dentro de la cotidianidad de la Universidad. En el desarrollo de la fase II se abordará el perfeccionamiento del modelado y diseño de los datos de la entidad PERSONA, la cual existe en los diferentes sistemas de la información corporativos de la Universidad como respuesta a los objetivos de cada macroproceso indicados en la **Tabla 14**.

Tabla 14 Macroprocesos, procesos y sistemas de la información enmarcados en el alcance de la fase I para la dimensión de modelado y diseño de datos

MACROPROCESO	PROCESO/ ETAPA	SISTEMA DE INFORMACIÓN
Misionales – Formación	Admisiones	Sistema de la información de Admisiones (ISYS)

Misionales – Formación	Gestión de la Actividad Académica	Sistema de la información Académica (SIA)
Misionales – Formación	Registro y Matrícula	Sistema de la información Académica (SIA)
Misionales – Investigación y Creación Artística	Gestión de la Investigación y Creación Artística	Sistema de la información (HERMES) componente de investigación
Misionales – Extensión	Gestión de la Extensión	Sistema de la información (HERMES) componente de extensión
Apoyo – Bienestar Universitario	Bienestar Universitario	Sistema de la información de Bienestar Universitario (SIBU)
Apoyo – Gestión de Talento Humano	Gestión de Talento Humano	Sistema de la información de Talento Humano (SARA)

El modelo de datos físico es la base de la implementación de una solución de hardware, software y comunicaciones que responde al modelo lógico. En el caso de la Universidad Nacional, las funciones de la sección de aplicaciones de la Dirección Nacional de Estrategia Digital (CSU, Acuerdo 316, 2019) se definen en la Resolución de Rectoría 318 de 2020, que incluye la de: “diseñar y mantener el modelo de planeación de los sistemas de la información de la Universidad, incorporando aspectos relacionados con catálogo y directorio de sistemas de la información, arquitectura de sistemas de información, propiedad intelectual, entre otros” (Numeral 1, Artículo 7).

Además, se establece la función de “Desarrollar y mantener los esquemas de integración de los sistemas de la información de la Universidad, incluyendo la integración con sistemas de externos” (Numeral 4, Artículo 7). Dado que la puesta en marcha de la sección de aplicaciones es reciente en la institución, se considera pertinente incluir los modelos de datos físicos en una fase posterior y de manera articulada con la Dirección Nacional de Estrategia Digital.

4.3. Desarrollo de la fase II de modelado y diseño de datos

Tabla 15 SIPOC de la dimensión de modelado y diseño de datos

Proveedores	Entradas	Actividades	Salidas	Usuarios
Referentes de modelado de datos, macroprocesos de la universidad	Modelo de GI, diagnóstico de problemas identificados en sistemas de información	Implementar metodología de identificación y documentación de los elementos para el modelamiento	Modelo conceptual de datos, modelo lógico de datos	DNED, Arquitectos y desarrolladores de aplicaciones
DNED	Diagnóstico de problemas identificados en sistemas de información	Establecer instrumentos de documentación de elementos para el modelamiento	Formulario de recolección y validación de datos e información, Formato de documentación de datos e información	Administradores de bases de datos, gestores de productos y servicios de información
DNED	Modelo conceptual (versión 1.0), modelo	Diagramar y socializar modelos de datos	Modelo conceptual (versión 2.0), modelo	Administradores de bases de datos, gestores de productos y servicios

lógico de datos (versión 1.0)	lógico de datos (versión 2.0)	de información, equipo técnico y funcional de los sistemas de la información existentes
-------------------------------	-------------------------------	---

4.3.1. Proveedores

Para el desarrollo de la dimensión se requiere, por una parte, un proceso de actualización permanente en cuanto al modelamiento de datos, teniendo en cuenta que la gestión de la información es una rama del conocimiento en permanente evolución. Por otra parte, se requiere integrar esfuerzos con los profesionales miembros de las dependencias que tienen bajo su responsabilidad la parte técnica y funcional de los sistemas de la información de la Universidad o que actúan como interlocutores funcionales para el soporte técnico de las soluciones que se tienen tercerizadas a un proveedor. Son ellos quienes manejan un vocabulario común alrededor de los datos (DAMA International, 2017), tienen un conocimiento explícito acerca de la operación de los sistemas de la información y dan respuesta a los requerimientos de la información interna y externa a la Universidad de acuerdo con los procesos a los que da apoyo cada dependencia.

4.3.2. Entradas

Establecer el concepto de los datos requiere de entender todos los procesos y procedimientos que se llevan a cabo al interior de la Universidad para dar cumplimiento a sus fines misionales, para esto deben analizarse los siguientes insumos:

- Documentación de los procesos de la Universidad.
- Instructivos, guías y procedimientos de las actividades que se realizan en los sistemas de la información de acuerdo con el alcance establecido para la fase I.
- Experiencia de los equipos técnicos y funcionales de los sistemas de la información que soportan procesos misionales, de acuerdo con el alcance establecido para la fase I.
- Observación directa sobre los sistemas de información.

Con el fin de identificar requerimientos para fortalecer la operación y gestión de la Universidad, además de enfocar una estructura que obedezca a un análisis de arquitectura de datos, es importante contar con los siguientes insumos:

- Diagnóstico de las necesidades de la información (Universidad Nacional de Colombia - Grupo de Gestión de Información, 2021) donde se identifican requerimientos no cubiertos y aspectos a fortalecer o mejorar de las necesidades satisfechas.
- Diagnósticos de problemas identificados por debilidades en la definición de los datos que se gestionan en la Universidad.
- Alcance de la dimensión de arquitectura (ver sección 3.2.2 en p. 55) en cuanto a la forma en que deben estructurarse los datos de acuerdo con la dinámica de funcionamiento propia de la Universidad.

4.3.3. Actividades

4.3.3.1. Implementar metodología de identificación y documentación de los elementos para el modelamiento

4.3.3.1.1. Identificar requerimientos para realizar el modelamiento

Para generar los modelos conceptual y lógico se requiere identificar los siguientes elementos con los cuales se analizan los sistemas de la información corporativos y se establece la base para la construcción de modelos de nuevos sistemas de información:

- Sistema de la información donde se llevan a cabo las transacciones.
- Entidades que intervienen en la transacción.
- Definición y marco normativo de la transacción.
- Atributos que hacen parte de la transacción.
- Periodicidad con la que se lleva a cabo la transacción.
- Identificadores o llaves que intervienen en la transacción.

4.3.3.1.2. Identificar una transacción a modelar de acuerdo con su marco normativo

Cada una de las dependencias encargadas de la administración de los sistemas de la información o de los desarrollos de nuevas aplicaciones basa las operaciones que se realizan en cada sistema en los marcos normativos que definen las actividades para el cumplimiento de los objetivos misionales. De esta forma, un análisis sobre los marcos normativos permite identificar las posibles transacciones que deben modelarse, donde en muchos casos y dada la naturaleza de la Universidad, se puede ofrecer una especificidad suficiente para una exploración inicial de atributos partiendo de los requerimientos funcionales. Además, desde los sistemas de la información existentes, se dispone de los manuales de operación, instructivos, guías y procedimientos que también aportan una visión bastante acertada para identificar los datos que pueden encontrarse en cada uno de ellos.

Adicionalmente, en el diagnóstico de necesidades (Grupo de Gestión de Información, 2021) pueden encontrarse necesidades no cubiertas que deberán abordarse según el análisis de la existencia de los datos, por lo cual los requerimientos que pretendan responderse serán la línea base del modelamiento.

Partiendo de los anteriores insumos se establece una transacción que puede darse en un sistema, bien sea el punto de inicio de una entidad o las relaciones que se deriven entre entidades, donde se tenga un entendimiento claro del requerimiento funcional que se quiere operacionalizar y desde allí se determine la meta de modelar dicha transacción.

4.3.3.1.3. Identificar las entidades que intervienen en la transacción

Las entidades son objetos del mundo real (persona, organización, lugar, cosa), sobre las que una organización recaba información. Las entidades usualmente se nombran mediante sustantivos y tienen

atributos propios. Para el objetivo del modelamiento de datos se debe analizar cuáles entidades aportan atributos a la transacción. Generalmente se caracterizan porque sus llaves son incluidas en la transacción y evidencian la existencia de relaciones o punto de inicio de esa entidad.

Un ejemplo de esta identificación es la transacción para abrir una historia académica para un estudiante. En dicha transacción intervienen las llaves de identificación de los planes de estudio (código único por cada malla curricular) y las llaves de identificación de los estudiantes (tipo y número de documento), lo que genera combinaciones únicas que definen la asignación de una nueva historia académica. Por lo tanto, las entidades que intervienen son: PLAN DE ESTUDIO y PERSONA.

4.3.3.1.4. Establecer los atributos que definen cada entidad, modelando sus puntos de inicio

En aquellas situaciones donde se identifiquen transacciones de registro o actualización de los atributos que definen una entidad se hace necesario crear un punto de inicio. Para ello se establecen todos los atributos propios de la entidad; por ejemplo: la transacción de crear o actualizar el registro de personas en un sistema de la información implica un punto de inicio para la entidad PERSONA, donde se incluyen los atributos de tipo y número de identificación, nombres, apellidos, direcciones, teléfonos, entre otros. Los atributos serán los disponibles en el sistema de la información corporativo dado el alcance definido para esta fase; sin embargo, debe revisarse la pertinencia de agregar y actualizar el diseño de nuevos atributos que contribuyan a la definición de la entidad.

Los insumos de identificación de los atributos existentes en esta fase se encuentran en:

- Los manuales e instructivos de operación de los sistemas de información.
- La normativa interna y externa que define varias entidades.
- Observación directa del sistema de información.
- Aporte de los líderes funcionales de cada uno de los sistemas de información.

4.3.3.1.5. Establecer los atributos que hacen parte de la transacción a modelar como relaciones entre entidades

En aquellas situaciones donde se evidencie que dos o más entidades intervienen en la transacción, deben establecerse relaciones. Para ello es importante identificar las llaves de cada una de las entidades que intervienen y adicionar los atributos que se generan como producto de la transacción. Por ejemplo: la transacción de crear la historia académica de un estudiante implica la participación de las entidades PERSONA y PLAN DE ESTUDIO, ya que se identifica que el código del plan de estudios (llave de la entidad PLAN DE ESTUDIOS) y el tipo y número del documento de identidad del estudiante (llaves de la entidad PERSONA) intervienen en la transacción, además, se generan atributos adicionales como el número de la historia académica, entre otros.

Los atributos en este caso responden a los identificados como existentes, dado el alcance de esta fase, pero en el caso de nuevas aplicaciones, debe orientarse hacia lo necesario para cumplir con la operación y el funcionamiento de nuevos requerimientos. Esto orientado a que intervenga más de una entidad. Los insumos son los mismos que los indicados en el numeral inmediatamente anterior (4.3.3.1.4).

4.3.3.2. Establecer instrumentos de documentación de elementos para el modelamiento

Una vez se realiza el levantamiento de los elementos pertinentes para el modelo, se procede a documentar y validar con el equipo técnico y funcional elementos como codificaciones, llaves y nombres de los atributos a modelar. Este proceso se lleva a cabo mediante mesas de trabajo y por medio de la documentación en el **Formulario para el levantamiento de datos e información (Anexo 5)**.

4.3.3.3. Diagramar y socializar modelos de datos

4.3.3.3.1. Diagramar el modelo conceptual y lógico

Posterior a la validación con los equipos se procede a graficar de acuerdo con los estándares aplicables tanto el modelo conceptual como el modelo lógico. De esta forma se documentan de una forma holística los datos gestionados en la Universidad de acuerdo con el alcance establecido para esta fase.

4.3.3.3.2. Socializar los diagramas con proveedores y usuarios de la dimensión

Los diagramas del modelo conceptual y del modelo lógico deberán ser una fuente de consulta permanente para los miembros de los equipos técnicos y funcionales de los sistemas de información, especialmente quienes tienen un papel activo en la gestión de información. Por ello deberán estar contenidos en el repositorio de metadatos de negocio y generar la cultura de consulta ante requerimientos de información, la generación de nuevas aplicaciones o el análisis de intercambios de información. En este sentido se pueden utilizar instrumentos de documentación posterior al levantamiento. Esta es una actividad que deberá iniciar a desarrollarse conforme se avance en las iniciativas de transformación digital de la Universidad.

4.3.3.3.3. Actualizar permanentemente los modelos

Como se ha mencionado anteriormente, los modelos representan una visión holística de los datos de la Universidad, por lo tanto, (y debido a que cada sistema de la información tiene políticas y dinámicas propias de actualización, mantenimiento y desarrollo) es importante que se cree la cultura de actualización de los modelos, de forma que se convierta en un insumo fundamental para la gestión de información.

4.3.4. Salidas

Los artefactos producidos anteriormente para esta dimensión se diseñaron, por una parte, con el objetivo de recoger y documentar los insumos necesarios para los modelos y, por otra parte, para construir y entregar un prototipo del modelo conceptual (**Anexo 6**) y el modelo lógico (**Anexo 7**). Estos diseños deben ser perfeccionados y ajustados en la fase II. Todos los artefactos fueron aplicados. En esta dimensión se relacionan los siguientes anexos:

- Anexo 5: Formulario para el levantamiento de datos e información.

- Anexo 5B: Instructivo de diligenciamiento para el formulario de levantamiento de datos e información.
- Anexo 5C: Propósito de las secciones del formulario para el levantamiento de datos e información.
- Anexo 6: Prototipo de modelo conceptual.
- Anexo 7: Prototipo de modelo lógico.
- Anexo 8: Instrumento de documentación de datos e información.

4.3.5. Usuarios

Los usuarios de los modelos conceptual y lógico son aquellos que tienen un panorama sobre los datos que existen y se gestionan en la Universidad, de forma que puedan apoyar sus decisiones para:

- Desarrollar nuevas aplicaciones, dado que se tiene el panorama de oportunidades de integración y aporte a requerimientos de datos no cubiertos en la actualidad.
- Administrar bases de datos, debido a que se pueden identificar de forma holística todas las instancias de bases de datos y su contenido.
- Gestionar productos y servicios de información, ya que se tiene una vista holística de los datos se pueden establecer alcances en la creación de productos y servicios de la información a la vez que puede analizarse requerimientos para identificar el origen de los datos de respuesta o informar cuando no se cuenta con los datos requeridos.

Los equipos técnicos y funcionales son usuarios de los modelos conceptual y lógico. Estos modelos son herramientas que les permiten identificar los datos y su relación con otros sistemas de información, así como hacer análisis e identificar oportunidades de mejora.

5. INTEGRACIÓN E INTEROPERABILIDAD DE DATOS

5.1. Conceptualización

Teniendo en cuenta tanto la diversidad de la información como la diversidad de los sistemas que la administran al interior de la Universidad, la integración e interoperabilidad son componentes esenciales en la gestión de información. La integración e interoperabilidad es la habilidad de que varios sistemas se comuniquen (DAMA Internacional, 2017), por lo que describen procesos relativos al movimiento y la consolidación de datos dentro y entre almacenes de datos, aplicaciones u organizaciones. La integración consolida los datos en formas coherentes, ya sean físicas o virtuales.

Esta dimensión depende de otras dimensiones de la gestión de datos como:

- **Gobierno de datos:** para definir los roles de interoperabilidad.
- **Arquitectura de datos:** para ubicarse dentro del diseño global de arquitectura y obtener los descriptores y nomenclatura de componentes de información.
- **Seguridad de datos:** para garantizar que las soluciones protejan apropiadamente la seguridad de los datos, ya sean persistentes, virtuales o que se encuentren en movimiento entre aplicaciones y organizaciones.
- **Metadatos:** para el seguimiento del inventario técnico de datos (persistentes, virtuales y en movimiento), el significado de negocio de los datos, las reglas de negocio para transformar los datos, así como su historia operacional y el linaje de los datos.
- **Almacenamiento de datos y operaciones:** para gestionar las instancias físicas de las soluciones.
- **Modelo y diseño de datos:** para diseñar las estructuras de datos incluyendo la persistencia física en bases de datos, estructuras de datos virtuales y mensajes que pasan información entre aplicaciones y organizaciones.

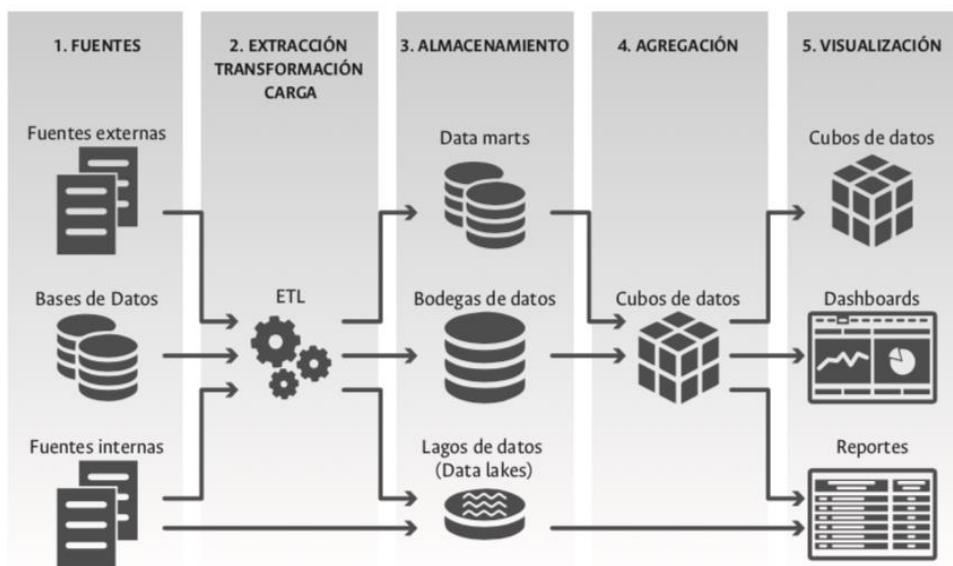
El modelo DAMA ofrece un estándar internacionalmente aceptado aplicable en la práctica, recopilando experiencias y análisis sobre implementaciones en diferente tipo de organizaciones, llegando a consolidar un conjunto de buenas prácticas aplicables a los datos, siendo de esta manera un referente adecuado (similar al *Project Management Institute* con su PMBOK® en el caso de la gestión de proyectos).

Adicionalmente, es necesario tener en cuenta el contexto y la regulación local, en particular sobre las entidades del sector público, por lo cual se adoptan elementos de la Metodología IT4+ del MinTIC (2019), en particular el modelo de interoperabilidad y la definición conceptual del lenguaje común de intercambio, buscando afinar detalles específicos de implementación.

5.1.1. Integración e interoperabilidad de datos: definición

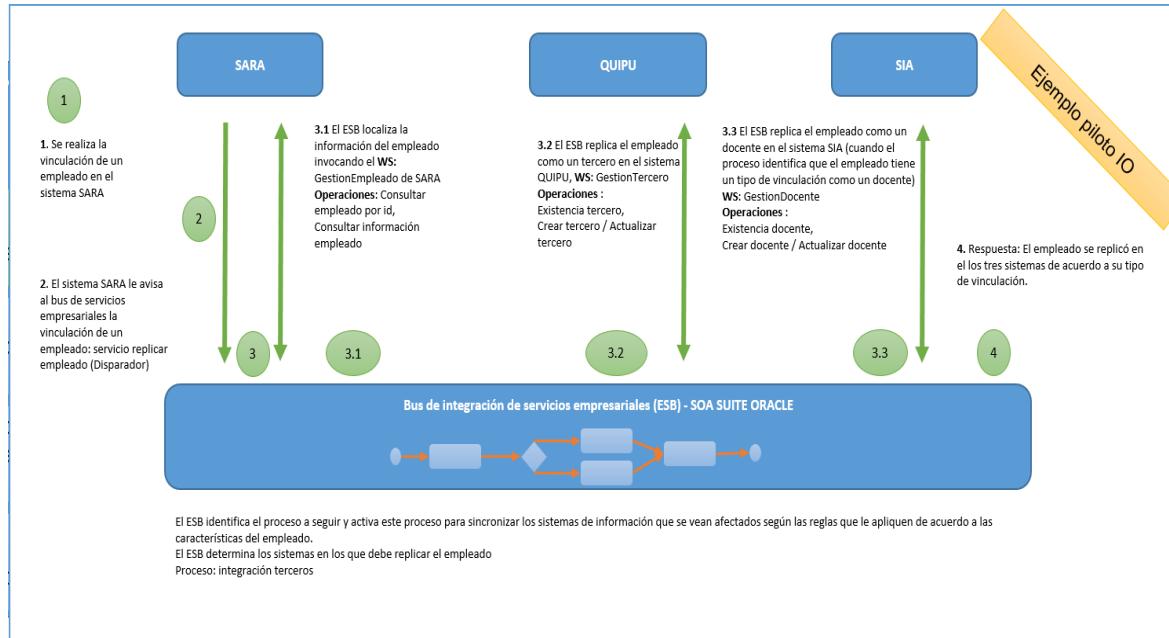
La integración de datos es el intercambio unidireccional de datos a partir de una o varias fuentes (extracción) hacia uno o varios destinos realizando en el camino una transformación de esos datos (que puede ser estandarización, consolidación, verificación de métricas de calidad, generación de registros con errores, entre otras) y su carga en un destino de la información apropiado para realizar tanto reportes estandarizados como consultas analíticas. La integración está usualmente relacionada con la construcción de bodegas de datos (*Datawarehouse - DWH*) y su ámbito de aplicación son los conjuntos masivos de datos. Por ejemplo: carga de admitidos en el SIA o en una bodega de datos de integración. La **Figura 18** ilustra el concepto de integración ampliamente usado en diversas compañías del sector público y privado.

Figura 18 Arquitectura tecnológica típica de un modelo de inteligencia de negocios – integración.
Fuente: Rodríguez y Bernal (2019)



La interoperabilidad es el intercambio bidireccional de datos de una o varias fuentes hacia uno o varios destinos donde se usa un lenguaje de comunicación previamente definido. Tiene como mesa de trabajo un entorno común de acceso donde se publica y se consume cierta información específica definida por las áreas de negocio. Por lo tanto, la interoperabilidad de datos está directamente relacionada con la construcción de un Bus Empresarial de Servicios (*Enterprise Service Bus - ESB*) orientado a la interoperabilidad. Su ámbito de aplicación son registros o transacciones específicas de datos; por ejemplo, para un evento que tiene un número de cédula específico, traer los datos de los diferentes sistemas que corresponden a ese registro. Lo anterior también permite mantener sincronizados los sistemas transaccionales participantes. La **Figura 19** ilustra el concepto de interoperabilidad de datos en la Universidad.

Figura 19 Esquema de interoperabilidad de datos. Fuente:
<https://dntic.unal.edu.co/index.php/proyectos/interoperabilidad>



Los conceptos de Integración e Interoperabilidad son complementarios y están orientados a diferentes necesidades de la organización. Existen diferentes grados de intercambio de información, tanto al interior de las organizaciones (*interoperabilidad interna*), así como también hacia el medio que las rodea (*interoperabilidad externa*).

El proyecto de Gestión de la Información de la Universidad Nacional usa de manera complementaria otros referentes del término interoperabilidad, entre los que se encuentran:

- La Guía de implementación de la facilitación del comercio de las Naciones Unidas⁴, donde se determina que la interoperabilidad es la capacidad de dos o más sistemas o componentes para intercambiar información y usar la información que se ha intercambiado. La interoperabilidad típica se lleva a cabo en dos niveles: semántico y técnico. La interoperabilidad semántica permite a las organizaciones o partes involucradas elaborar bibliotecas donde la información se presenta de una manera consistente, sin importar la aplicación o plataforma tecnológica usada para tal fin. La interoperabilidad técnica intercambia los datos a través de un conjunto común de formatos de intercambio para leer y escribir los mismos formatos de archivo y para usar los mismos protocolos.

⁴ <https://tfig.unece.org/SP/index.html>

- El vocabulario de la información y tecnología ISO/IEC 2382, donde se define la interoperabilidad como “la capacidad de comunicar, ejecutar programas o transferir datos entre varias unidades funcionales de manera que el usuario no tenga que conocer las características únicas de estas unidades” (ISO, 2015).
- El Marco de interoperabilidad propuesto por MinTIC - Gobierno Digital, según el cual “la interoperabilidad permite fortalecer la visión de unidad del Estado, al tener una mayor capacidad de comunicación, entrega y uso de servicios digitales de valor para mejorar la calidad de vida de los ciudadanos” (MinTIC, 2019, p. 8).

La Universidad puede hacer uso y aprovechamiento de los lineamientos establecidos en este marco para cada una de sus implementaciones, según sea el caso. En ese sentido, la interoperabilidad considera múltiples aspectos o dominios; no obstante, dentro de este marco de gobierno digital del MINTIC se trabaja la interoperabilidad en cuatro dominios específicamente: político, organizacional, semántico, técnico; los cuales son visibles en el Modelo conceptual del marco de interoperabilidad MinTIC. En la **Figura 20** se representan dichos dominios y, a continuación, se describe en detalle cada uno de ellos.

Figura 20 Modelo conceptual del marco de interoperabilidad MinTIC. Fuente:
<https://www.mintic.gov.co/arquitecturati/630/w3-propertyvalue-8117.html>



- **Dominio político:** corresponde a la disposición de un conjunto de políticas y normas que permiten el intercambio de información. La interoperabilidad político-legal consiste en garantizar que las entidades públicas realicen el intercambio de la información ajustado al marco jurídico vigente, que las políticas y estrategias puedan trabajar juntas y no se obstaculice o impida la interoperabilidad. Este dominio facilita la interoperabilidad en los posteriores dominios semántico y técnico de acuerdo con los lineamientos establecidos; así mismo

establece actividades, responsables y artefactos que ayuden a la entidad para determinar el grado de madurez en este dominio.

- **Dominio organizacional:** define los objetivos de la entidad, modela los procesos y facilita la colaboración de administraciones que desean intercambiar información y pueden tener diferentes estructuras organizacionales y procesos internos. Además de eso, busca orientar, con base en los requerimientos de la comunidad usuaria, los servicios que deben estar disponibles, fácilmente identificables, accesibles y orientados al usuario. Así mismo, permite construir claramente la relación que se da entre las entidades cuando son proveedoras o consumidoras de la información intercambiada y formaliza la relación mutua, sus alcances y responsabilidades frente a su rol en el proceso de interoperabilidad.
- **Dominio semántico:** permite garantizar que, en el momento de intercambiar datos, el significado de la información sea exacto y el mismo para todas las partes interesadas. De igual manera, permite que las entidades del Estado colombiano puedan estandarizar, gestionar y administrar su información. En este dominio semántico se consolida la arquitectura de interoperabilidad de datos que está diseñada para definir, clasificar y gestionar la información, así como los datos que conformarán el estándar.
- **Dominio Técnico:** hace referencia a las aplicaciones e infraestructuras que conectan sistemas de la información a través de los servicios de intercambio de información. Concretamente: interfaces, protocolos de interconexión, servicios de integración de datos, presentación e intercambio de datos y protocolos de comunicación seguros.

El modelo conceptual del MinTIC tiene ciertos principios fundamentales: enfoque al ciudadano, cobertura y proporcionalidad, seguridad, protección y reservación de la información, colaboración y participación, simplicidad, neutralidad tecnológica y adaptabilidad, reutilización, confianza y costoefectividad descritos detalladamente en su marco de interoperabilidad (2019c).

5.1.2. Integración e interoperabilidad de datos: elementos

En el marco metodológico del DAMA, se identifica la cadena de valor de la dimensión de integración e Interoperabilidad tanto a nivel de artefactos de entrada, actividades y salidas, así como a nivel de actores y partes interesadas como los proveedores, participantes y consumidores de dichos artefactos, donde son especialmente relevantes los estándares de datos que se vayan a aplicar. Los artefactos que son fundamentales en este contexto para la Integración e Interoperabilidad son: arquitectura de interoperabilidad, especificaciones de intercambio de datos, acuerdos de acceso a datos, servicios de intercambio de datos. DAMA (2017) constituye la línea base a nivel metodológico para su aplicación en el contexto particular de la Universidad Nacional de Colombia.

5.1.3. Integración e interoperabilidad de datos: estándar del lenguaje común de intercambio

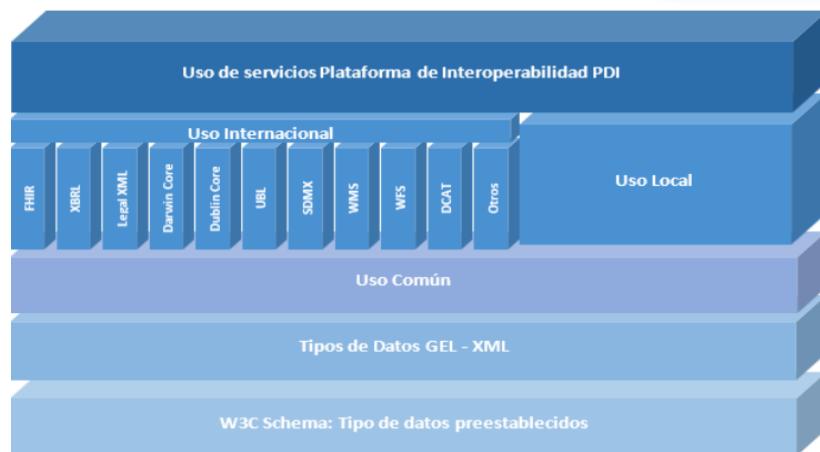
El intercambio de la información entre entidades del Estado requiere de acuerdos para definir las estructuras a partir de las cuales se intercambiarán los datos requeridos. Este proceso se hace dispendioso si es necesario llegar a acuerdos entre las diferentes entidades, cada una en procesos independientes. Para facilitar la definición de estas estructuras en el contexto local, MinTIC definió el Lenguaje Común de Intercambio de información, el cual brinda un significado y una estructura unificada sobre los datos, facilitando el entendimiento del negocio y el intercambio de la información de la organización, facilitando su gestión y su relación con el ciudadano. Así mismo puede ser utilizado por las entidades para obtener la información de los ciudadanos, a través de formularios, o para entregar información estructurada.

El lenguaje común se caracteriza por abordar la información que se usa o intercambia desde el punto de vista conceptual y técnico. Además, dentro del proceso de intercambio de información, el lenguaje común permite unificar el significado y la estructura de los conceptos a intercambiar, evitando que un mismo concepto tenga diferentes interpretaciones. El Lenguaje Común de Intercambio de la información también puede ser utilizado para proveer y/o consumir servicios de intercambio de información, estructurar formularios para la obtención de la información digital, y/o definir las estructuras para almacenamiento de la información.

Debido a que el *elemento de dato* se define como la unidad básica de información, éste identifica y representa un concepto a intercambiar, teniendo en cuenta sus relaciones, validaciones y estructura. En ese caso un *elemento de dato* puede ser de dos tipos: simple y/o compuesto.

Para organizar los *elementos de dato*, el lenguaje común de Intercambio de la información cuenta con una arquitectura de interoperabilidad de datos que se basa en capas. Cada una de las capas incluye los *elementos de dato* que aplican en cada capa respectiva.

Figura 21 Diagrama de arquitectura de interoperabilidad de datos del estándar de integración de datos e interoperabilidad. Fuente: <http://lenguaje.mintic.gov.co/aprende-del-lenguaje>



Los *elementos de dato* y sus clasificaciones se encuentran disponibles en el diccionario de elementos de datos⁵ y se clasifican de acuerdo con su uso y función de la siguiente manera:

- **Uso común:** conceptos que tienen la misma interpretación o significado, independientes de la localización geográfica o proceso de intercambio de la información en el que se utilicen.
- **Uso local:** conceptos que tienen la misma interpretación o significado, independientes de la entidad, trámite, servicio o proceso de intercambio de la información en el que se utilicen a nivel nacional.
- **Uso sectorial:** conceptos que tienen una interpretación particular o son de uso específico en un sector.
- **Uso servicios de la Plataforma de Interoperabilidad (PDI):** conceptos que son de uso específico de la Plataforma de Interoperabilidad (dominio técnico).
- **Uso internacional:** concepto que tiene *elementos del dato* que hacen parte o se encuentran definidos en estándares internacionales y que cubren las necesidades del lenguaje común de intercambio de información.

Existen varias organizaciones del país que han realizado implementaciones del proceso de estandarización del lenguaje de interoperabilidad y tienen actualmente publicados los servicios de intercambio de información. La incorporación del estándar implica la adopción del Marco de Interoperabilidad. Dicha incorporación puede definirse como un proceso progresivo basado en tres niveles donde la organización deberá cumplir una serie de pasos que le otorgan certificaciones en cada nivel.

5.2. Propósito

La gestión de la información en la Universidad ha venido creciendo de acuerdo con las necesidades específicas de cada área, lo cual ha generado desarrollos independientes que de manera individual cumplen adecuadamente su función. Sin embargo, ese desarrollo por silos presenta un reto importante para comunicar los sistemas de la información entre sí, dado que manejan conceptos, estándares y valores diferentes. Solucionar esta situación es el propósito de la dimensión de integración e interoperabilidad.

5.2.1. Metas

- Identificar las necesidades de intercambio de la información en los dominios organizacional y semántico, teniendo en cuenta el ciclo de vida del dato.
- Articular el proyecto de Interoperabilidad con la Política de Gestión de la Información.
- Proponer un procedimiento de gestión de elementos de interoperabilidad, desde las preguntas iniciales hasta la puesta en marcha, incluyendo los estándares a nivel semántico que aplicarán.

⁵ Disponible para consulta en: <http://lenguaje.mintic.gov.co/diccionario-de-elementos-de-datos>

- Recopilar y generar un esquema de construcción de los servicios de intercambio de datos estructurados más relevantes y los acuerdos de servicio asociados.

5.2.2. Alcance en la fase II

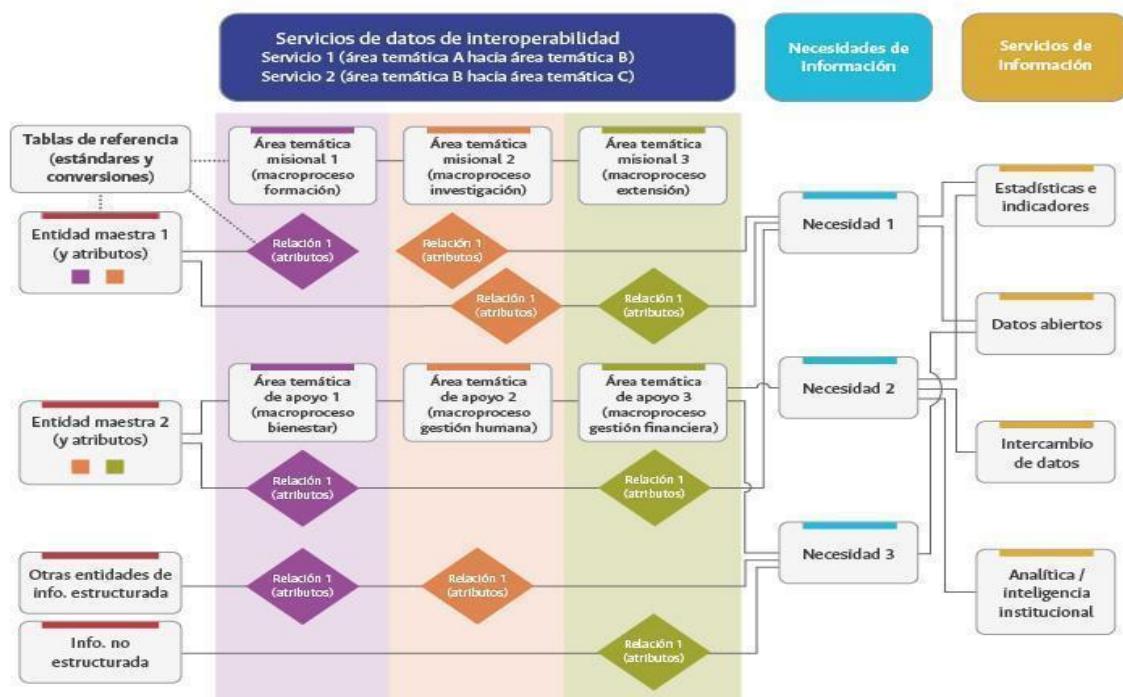
Teniendo en cuenta los 4 dominios de la integración e interoperabilidad existentes (político, organizacional, semántico y técnico), en esta fase del proyecto se hará énfasis en los dominios organizacional y semántico. La Interoperabilidad en su dominio técnico está más allá del alcance de esta fase del proyecto, ya que corresponde a las decisiones de implementación por parte de las áreas de tecnología de la Universidad. De igual manera, en la segunda fase se tendrán en cuenta específicamente los macroprocesos mencionados a nivel global del Modelo de Gestión de la Información.

5.2.3. Punto de referencia para la interoperabilidad

Respecto al diseño de arquitectura de datos, la interoperabilidad se ubica como una capa transversal a los diferentes macroprocesos. Como se muestra en la **Figura 22**, los servicios de datos que hacen parte de la integración e interoperabilidad a nivel interno se identifican en color violeta. Por otro lado, en la parte derecha se puede observar un servicio de intercambio de datos, donde participa la integración e interoperabilidad externa.

Para este efecto, es conveniente tener en cuenta: 1) los macroprocesos/procesos misionales implicados en el intercambio de información, 2) el detalle de la información que participa en el intercambio y 3) la notación y estándares para el intercambio de dicha información.

Figura 22 Interoperabilidad dentro del diseño de arquitectura de datos



5.3. Desarrollo de la fase II de integración e interoperabilidad de datos

Como se ha mencionado en la sección del alcance, el desarrollo de esta dimensión se enfocó en los dominios organizacional y semántico haciendo énfasis en las necesidades de intercambio de información, la definición de un lenguaje común de intercambio y la descripción de los principales servicios de información. Para poder unificar dichos conceptos, se buscó que las definiciones fueran preferiblemente comunes a la interoperabilidad interna (proveniente del proyecto de interoperabilidad) y la interoperabilidad externa (provenientes de los reglamentos del MinTIC) para así facilitar la comunicación.

El diagrama SIPOC que se presenta a continuación usa como referente de actividades el Marco de Interoperabilidad del MinTIC (2019c) en lo referente a los diagramas de flujo para los dominios organizacional y semántico. Teniendo en cuenta la necesidad de realizar intercambio de datos con otras entidades de una manera estandarizada, se siguió el estándar del lenguaje común de intercambio presentado en los antecedentes. Este lenguaje, además de elementos útiles a nivel organizacional y semántico, presenta en el dominio técnico el uso de Web Services basados en XML, lo cual es compatible con el trabajo realizado a nivel interno en el proyecto de interoperabilidad. No obstante, el lenguaje de implementación actual del proyecto de interoperabilidad (BPEL) requiere ciertos ajustes para su compatibilidad con el modelo que está adoptando el estado colombiano (XROAD).

Tabla 16 SIPOC para Integración e Interoperabilidad de datos en la UNAL

Proveedores	Entradas	Actividades	Salidas	Usuarios
DNED, Gerentes - directores de cada área temática o Macroproceso.	Objetivo y estrategias de negocio	Asignar responsables del intercambio de la información IOP	Identificación de responsables	DNED
DNED, Proveedores de la información - Áreas funcionales	Arquitectura de datos	Identificar los procesos susceptibles a interoperar	Identificación de procesos a interoperar	DNED
DNED, Proveedores de la información - Áreas funcionales	Necesidades y estándares de datos	Definir los requerimientos de integración e interoperabilidad y el ciclo de vida de los datos	Acuerdos de acceso a datos	DNED, Gerentes y ejecutivos
Expertos - Líderes funcionales de cada área temática o Macroprocesos	Necesidades y estándares de datos	Identificar oportunidades de mejora en los procesos implementando servicios de intercambio de información	Especificaciones de intercambio de datos (detalles del modo de intercambio)	DNED
DNED	Acuerdos de acceso a datos, Especificaciones de intercambio de datos (detalles del modo de intercambio)	Definir servicios de intercambio de la información Proveedor/Consumidor	Arquitectura de integración e interoperabilidad	Áreas temáticas y sistemas consumidores de la información de intercambio, Gerentes y ejecutivos
DNED	Arquitectura de integración e interoperabilidad	Divulgar los servicios de intercambio de la información al interior de la entidad	Servicios de la información socializados	Áreas temáticas y sistemas consumidores de la información de intercambio, Gerentes y ejecutivos
DNED	Servicios de la información socializados	Publicar servicios de intercambio de la información en el Directorio de Servicios	Directorio de servicios de la información actualizado	Áreas temáticas y sistemas consumidores de la información de intercambio, Entidades externas

5.3.1. Proveedores

- **Productores de datos (áreas funcionales encargadas de los procesos en la Universidad).** Para el macroproceso de Formación, la Dirección Nacional de la información Académica; para el macroproceso de Bienestar, la Dirección Nacional de Bienestar Universitario.
- **Comité directivo de TI (futuro grupo de gobierno de la información).** Comité de Estrategia Digital (Resolución 614 de 2020).
- **Ejecutivos y Gerentes (directivos de cada área temática / macroproceso).** Vicerrectores generales de investigación, académico y de sedes, directores de sede y directores de área del tema específico. director nacional de planeación, gerente financiero, secretaría general, director de TI.

- **Expertos de los temas de negocio y Analistas de datos (líderes funcionales y asesores de las áreas misionales).** En el macroproceso de Investigación y Extensión corresponde a la Vicerrectoría de Investigación y Extensión. Para el macroproceso de Formación, la dependencia es Vicerrectoría Académica y para los macroprocesos de Bienestar y Talento Humano la dependencia es la Vicerrectoría General. Para el macroproceso de Bienestar, la Dirección Nacional de Bienestar Universitario y finalmente, para el macroproceso de Gerencia Financiera y Administrativa es la rectoría junto con la Gerencia Financiera y Administrativa.
- Entidades externas proveedoras de información.

5.3.2. Entradas

- Objetivos y estrategias de negocio (Macroprocesos y procesos de la Universidad).
- Necesidades y estándares de datos provenientes de las matrices de cruce de actividades y necesidades de intercambio de información.
- Requerimientos regulatorios, de cumplimiento y de seguridad (normativa).
- Arquitecturas de datos provenientes de la dimensión de arquitectura de datos.
- Semánticas de datos que hacen parte del lenguaje común de intercambio y del lenguaje implementado actualmente (BPEL).
- Datos origen.

5.3.3. Actividades

5.3.3.1. Asignar responsables del intercambio de la información de Interoperabilidad

Esta actividad permite que la organización establezca los responsables a nivel funcional y técnico de los servicios de intercambio. Aquí participan directamente los lineamientos provenientes del gobierno de datos. Entre los principales responsables se destacan:

- Coordinación funcional: Líderes de cada proceso que interopera.
- Coordinación técnica: Comité de interoperabilidad.

5.3.3.2. Identificar los procesos susceptibles a Interoperar

Siempre que surge la necesidad de intercambiar información, se busca resolver el interrogante ¿qué intercambiar? Esta actividad debe permitir identificar todos aquellos datos en los procesos que pueden dar cumplimiento a la necesidad mencionada.

Una vez definido el responsable de la gestión de los servicios de intercambio de información, la entidad debe identificar las áreas misionales encargadas de los procesos de negocio susceptibles de interoperabilidad. Además, definir los ajustes organizacionales y de los procesos involucrados realizando un diagnóstico de activos y/o fuentes de la información de la entidad.

En el caso de la Universidad, los procesos susceptibles a interoperar hacen parte de la matriz de levantamiento de necesidades de intercambio de la información (**Anexo 3**). En este trabajo se identificó un total de 214 eventos como candidatos. Estos pertenecen a los siguientes once macroprocesos (se resaltaron con negrilla los contemplados en la primera fase):

1. Formación.
2. Gestión de investigación y creación artística.
3. Extensión.
4. Bienestar Universitario.
5. Gestión del Talento Humano.
6. Gestión Administrativa y Financiera.
7. Gestión de laboratorios.
8. Seguridad social.
9. Relaciones Interinstitucionales.
10. Direccionamiento Institucional.
11. Gestión de recursos y servicios bibliotecarios.

5.3.3.3. Definir los requerimientos de integración e interoperabilidad y el ciclo de vida de los datos

Existen unos lineamientos base definidos desde el proyecto de interoperabilidad que es necesario tener en cuenta:

- La información no debe ir de un sistema a otro directamente sin una debida validación y control.
- La comunicación entre los sistemas de la información se debe realizar mediante un *middleware* que se encargará de la oportuna entrega de la información al sistema que haga la respectiva solicitud.
- Manejar un formato neutro tanto para la emisión como para la recepción de datos por intermedio del *middleware* basado en XML para tener homogeneidad en el intercambio de información.
- Permitir colas persistentes para apoyar procesos que requieren integridad de la información y alta disponibilidad.
- Definir estándares de seguridad para el tratamiento de la información para la respectiva autorización y autenticación.
- Para el intercambio de la información que no esté soportada sobre el bus de servicios debe tener en cuenta políticas de confidencialidad de la información, y la vigencia de este intercambio, para casos específicos de ocurrencias de periodicidad baja.
- Para intercambio mediante archivos planos se debe tener en cuenta que el medio de intercambio ya no debe ser correo electrónico ni por medios físicos, se hará mediante adaptadores que están incluidos en el bus de servicios.

5.3.3.4. Identificar oportunidades de mejora en los procesos implementando servicios de intercambio de la información

En conjunto con los responsables de la información del proceso se debe realizar un análisis que permita determinar las modificaciones necesarias para convertirlo en un servicio de intercambio de información.

5.3.3.5. Definir servicios de intercambio de la información proveedor/consumidor

Con base en el diagnóstico de flujos y activos de la información e identificada la necesidad de intercambio de la información la entidad deberá tener en cuenta los lineamientos descritos en el dominio técnico para la creación de los servicios.

5.3.3.6. Divulgar los servicios de intercambio de la información al interior de la entidad

La entidad debe interiorizar y apropiar los servicios de intercambio de la información desarrollados para mejorar sus procesos de negocio, a través del desarrollo de un plan de capacitación para generar capacidades en interoperabilidad para sus funcionarios.

5.3.3.7. Publicar servicios de intercambio de la información en el Directorio de Servicios

La entidad realizará el registro de los servicios de intercambio de la información que presta en el directorio general de servicios de intercambio de información.

5.3.4. Participantes y roles híbridos

- **Arquitecto de datos:** analista senior responsable de la arquitectura de datos y de la integración de datos. Los arquitectos de datos pueden trabajar a nivel empresarial o a nivel funcional y especializarse en *data warehousing*, *data marts* y sus procesos de integración asociados.
- **Modelador de Datos:** responsable de capturar y modelar los requerimientos de datos, definiciones de datos, reglas de negocio, requerimientos de calidad de datos y modelos de datos físicos y lógicos.
- **Administrador del Modelo de Datos:** responsable del control de versiones y del control de cambios del modelo de datos.
- **Administrador de Bases de Datos:** responsable del diseño, implementación y soporte de activos de datos estructurados y del rendimiento de la tecnología que hace que los datos sean accesibles.
- **Administrador de Seguridad de Datos:** responsable de garantizar el acceso controlado a los datos que requieren diferentes niveles de protección.
- **Arquitecto de Integración de Datos:** desarrollador senior de integración de datos responsable de diseñar tecnología para integrar y mejorar la calidad de los activos de datos empresariales.

- **Especialista en Integración de Datos:** diseñador de software o desarrollador responsable de implementar sistemas para integrar (replicar, extraer, transformar, cargar) activos de datos, en lotes o casi en tiempo real.
- **Desarrollador de Analítica/Reportes:** desarrollador de software responsable de crear soluciones y aplicaciones analíticas.
- **Arquitecto de Aplicaciones:** desarrollador senior responsable de integrar sistemas de aplicaciones.
- **Arquitecto Técnico:** ingeniero técnico senior responsable de coordinar e integrar la infraestructura de TI y el portafolio tecnológico de TI.
- **Ingeniero Técnico:** analista técnico senior responsable de investigar, implementar, gestionar y respaldar una parte de la infraestructura tecnológica de la información.
- **Administrador de la Mesa de Ayuda:** responsable de manejar, rastrear y resolver problemas relacionados con el uso de la información, los sistemas de la información o la infraestructura de TI.
- **Auditor de TI:** auditor interno o externo de las responsabilidades de TI, incluida la calidad de los datos y la seguridad de los datos.
- **Analista de Calidad de Datos:** responsable de determinar la idoneidad de los datos para su uso y supervisar el estado actual de los datos. Contribuye al análisis de causa raíz de los problemas de datos y ayuda a la organización a identificar los procesos del negocio y las mejoras técnicas que contribuyen a obtener datos de mayor calidad.
- **Especialista en Metadatos:** responsable de la integración, control y entrega de Metadatos, incluida la administración de repositorios de Metadatos
- **Arquitecto de Inteligencia de Negocio:** analista senior de inteligencia de negocio, responsable del diseño del ambiente de usuario de inteligencia de negocio.
- **Analista/Administrador de Inteligencia de Negocio:** responsable de apoyar el uso efectivo de los datos de inteligencia de negocio por parte de los profesionales del negocio.
- **Gerentes de programa de Inteligencia de Negocio:** coordina los requerimientos e iniciativas de BI en toda la corporación y los integra en un programa prioritario coherente y un mapa de ruta (*roadmap*).

5.3.4. Salidas

- **Arquitectura de elementos de dato e interoperabilidad:** Se adopta la arquitectura de elementos de dato presentada en los antecedentes definiendo las diferentes categorías de la **Figura 22**.

- **Necesidades de intercambio de datos:** Durante el proyecto de interoperabilidad mencionado en los antecedentes se llevó a cabo un levantamiento de necesidades de intercambio de datos a nivel interno donde se identificaron 214 eventos con sus correspondientes datos a intercambiar.

5.3.5. Usuarios

- Consumidores de la información (Ministerio de Educación Nacional, Vicerrectoría de Investigación, Dirección Nacional de Bienestar Universitario, Dirección Nacional de Innovación Académica, Gerencia Nacional Financiera y Administrativa).
- Trabajadores del conocimiento.
- Gerentes y ejecutivos.
- Entidades externas (Ministerio de Educación Nacional, Procuraduría General de la Nación).

6. GESTIÓN DE DATOS MAESTROS Y DE REFERENCIA

6.1. Conceptualización

6.1.1. Datos maestros

Los datos maestros definen los objetos de negocio centrales utilizados en diferentes aplicaciones de una institución. Junto a sus metadatos, atributos, definiciones, funciones, conexiones y taxonomías, los datos maestros representan los elementos más importantes de una organización; sin ellos se detienen los objetivos de negocio de la institución (Loshin, 2009).

6.1.2. Datos de referencia

Los datos de referencia son aquellos utilizados para caracterizar o clasificar otros datos. Hay varios niveles de complejidad entre ellos, desde los más simples (que relacionan un valor en forma de código con una descripción en lenguaje natural) hasta los más complejos (que incorporan mapeos y jerarquías). Los datos de referencia existen prácticamente en todos los almacenes de datos y pueden relacionar información externa o interna a una organización (DAMA Internacional, 2017).

6.1.3. Estructura de datos de referencia

En toda organización, dependiendo de la granularidad y complejidad con la que se representan los datos de referencia, éstos se pueden estructurar de diferentes maneras: desde listas simples, pasando por listas de referencia cruzada, hasta el máximo grado de complejidad presentándose en taxonomías (DAMA Internacional, 2017).

6.1.3.1. Lista simple

Una lista simple permite representar los datos de referencia mediante pares entre un valor en forma de código y una descripción (DAMA International, 2017). En la **Tabla 17** se presenta un ejemplo asociado al proceso de formación y a la manera en que el Sistema de la información Académica (SIA) maneja una lista simple. En ella, el valor del código hace referencia al identificador primario y la descripción es lo que representa el código. La descripción sirve para expresar objetos en lenguaje natural y el código se utiliza para realizar búsquedas y/o abreviaturas a la descripción asociada.

Tabla 17 Fragmento de lista de referencia simple. Ejemplo de códigos de bloqueo (SIA)

Valor del código	Descripción
20	Retiro por superar el tiempo máximo de permanencia permitido en el posgrado
26	Historia Académica inactiva por fallecimiento
31	Etapa inicial Sede de Presencia Nacional, SPN (Acuerdo 025 de 2007 CSU)

Dependiendo de la complejidad de los datos de referencia se pueden agregar atributos a la lista de referencia simple con el fin de proveer más información acerca del significado del código. Para algunos casos de listas de referencia simple en la Universidad se hizo necesario agregar algunos atributos como normativa referente a su definición o uso, caso de uso y fecha de última actualización.

6.1.3.2. Listas de referencia cruzada

Las listas de referencia cruzada también tienen una relación entre un código y una descripción, pero, a diferencia de las listas de referencia simple, en las listas de referencia cruzada se tiene equivalencia entre diferentes códigos que hacen referencia al mismo objeto. La **Tabla 18** muestra la lista de referencia cruzada entre códigos de sede para el Sistema de la información Académica (SIA) y el Sistema de la información de Admisiones (ISYS), además de la descripción que ambas representan.

Tabla 18 Lista de referencia cruzada. Ejemplo de equivalencia entre códigos de sede para SIA e ISYS

Valor código SIA	Valor código ISYS	Descripción SIA	Descripción ISYS
1101	1	Bogotá	Bogotá
1101	9	Bogotá	Sumapaz
1102	2	Medellín	Medellín
1102	11	Medellín	Sinifaná
1103	3	Manizales	Manizales
1103	10	Manizales	Caldas
1104	4	Palmira	Palmira
1104	12	Palmira	Choco
1124	5	Orinoquía	Orinoquía
1125	6	Amazonía	Amazonía
1126	7	Cariibe	Cariibe
9920	8	Tumaco	Tumaco
9933	13	La Paz	La Paz

6.1.3.3. Taxonomías

Las estructuras taxonómicas capturan la información en diferentes niveles de especificidad permitiendo clasificar el contenido para apoyar la inteligencia de negocio (DAMA, 2017). Este tipo de estructuras son evidentes en algunos sistemas de la información de la Universidad. Un ejemplo claro se da en el Sistema de la información Admisiones (ISYS), administrado por la Dirección Nacional de Admisiones, en donde las líneas de investigación de maestría siguen un nivel de jerarquía. De acuerdo con los ejemplos presentados en la **Tabla 19**, dado un programa padre (2115000 - Biociencias y derechos) se puede crear una nueva línea de investigación compuesta por un número único asociado al programa padre (2115013 asociado a 2115000) y un nombre compuesto por el programa padre y la línea de investigación correspondiente (Biociencias y derechos – Bioprospección y Biotecnología).

Tabla 19 Taxonomía. Ejemplo fragmento de líneas de investigación sistema de la información de admisiones ISYS

Código programa	Nombre de la línea de investigación	Nivel	Código de programa padre
2115000	Biociencias y derechos	-	2115000
2115013	Biociencias y derechos – Bioprospección y Biotecnología	Maestría	2115000
2115014	Biociencias y derechos – Bionegocios y Biocomercio	Maestría	2115000

6.1.4. Estándares del dominio de los datos

El dominio de los datos es un conjunto con todos los posibles valores que se pueden asignar a un atributo de la misma índole. Un ejemplo es la fecha: todos los atributos que referencian una fecha pueden, y generalmente toman, diferentes formatos, como: “aaaa/mm/dd” o “dd/mm/aaaa” por mencionar algunos. Sin importar su formato hacen parte del dominio de fecha. Ahora bien, es necesario estandarizar estos dominios para parametrizar la consolidación de los datos y así ayudar a mantener su calidad, fortaleciendo la gestión institucional de la información (para abordar este tema en mayor profundidad, se aconseja al lector remitirse al capítulo de gestión de metadatos, en el módulo de conceptualización).

6.2. Propósito

Gestionar datos de referencia en la Universidad tiene como propósito proveer herramientas para reconciliar diferencias conceptuales y de dominio de los datos que son comunes a varios de sus procesos o dependencias. Así mismo, aportar elementos para la estandarización desde los consumidores de datos que tengan interés en usar y explotar datos de la Universidad para diferentes fines. Finalmente, priorizar esfuerzos y actividades a partir de la definición de aquellos conjuntos de datos de gran importancia para la operación, gestión y cumplimiento de los fines misionales de la Universidad.

Gestionar datos maestros y de referencia, por un lado, contribuirá a mejorar los mecanismos de intercambio, integración e interoperabilidad ya que documenta y socializa todos los recursos de datos e información que gestiona y produce la Universidad y las diferentes interpretaciones y especificaciones de un mismo objeto. Por otro lado, establece la priorización de esfuerzos en la calidad, oportunidad y registro de estos datos, ya que se identifica su impacto en la gestión de la información de la institución. Finalmente, aporta al establecimiento de un lenguaje común sobre la comprensión de los datos existentes y la forma en que se encuentran dentro de cada sistema de información.

6.3. Metas

- Documentar y gestionar los datos maestros y de referencia para la Universidad con la terminología que se sustenta en sus marcos normativos o referentes externos aplicables.
- Documentar de forma explícita, y bajo una misma metodología, los datos maestros y de referencia por la Universidad, de acuerdo con el alcance establecido promoviendo instrumentos de trabajo colaborativo entre las dependencias líderes de fuentes de información.
- Identificar opciones de mejora en la estandarización de los datos de la Universidad a partir de los marcos normativos o referentes externos.

6.4. Alcance fase II

En el desarrollo de la fase II del Proyecto de Gestión de la información se abordarán los datos maestros y de referencia para:

- Establecer un concepto y una metodología de identificación de datos maestros y de referencia que permita a la Universidad priorizar los esfuerzos en su gestión.
- Proponer metodologías e instrumentos para documentar datos de referencia de tipo listas simples y listas de referencia cruzada.

Para las siguientes fases se propone abordar temas como las listas taxonómicas y la adquisición de datos externos para la limpieza, consistencia y mitigación de redundancia en esfuerzos. Un ejemplo de estas iniciativas sería la integración con la Registraduría Nacional para los datos básicos de las personas.

6.5. Desarrollo de la fase II de gestión de datos maestros y de referencia

Tabla 20 SIPOC de la dimensión de gestión de datos maestros y de referencia fase II

Proveedores	Entradas	Actividades	Salidas	Usuarios
DNED	Modelo de Gestión de la Información	Implementar la metodología de identificación de datos maestros y de referencia	Datos maestros y de referencia de la UNAL	Arquitectos y desarrolladores de aplicaciones, gestores de productos y servicios de información
DNED	Modelo de Gestión de la Información, definición de arquitectura de datos, diagnósticos sobre sistemas de información	Establecer instrumentos de documentación de datos maestros y de referencia	Propuesta de instrumento de documentación de datos maestros y de referencia	Arquitectos y desarrolladores de aplicaciones, gestores de productos y servicios de información
DNED	Datos maestros y de referencia de la UNAL	Incluir datos maestros y de referencia en modelos de datos y repositorio de datos	Modelos de datos y repositorio de datos actualizados	Equipo técnico y funcional de sistemas de información

6.3.1. Proveedores

Para el desarrollo de la dimensión se requiere, por una parte, un proceso de actualización permanente en cuanto a definiciones y gestión de datos maestros y de referencia, teniendo en cuenta que la gestión de la información es una rama del conocimiento en permanente evolución.

Por otra parte, se requiere integrar esfuerzos con los profesionales miembros de las dependencias que tienen bajo su responsabilidad la parte técnica y funcional de los Sistemas de la información corporativos de la Universidad o que actúan como interlocutores funcionales para el soporte técnico de las soluciones que se tienen tercerizadas a un proveedor ya que ellos conocen los datos mínimos requeridos para la operación de sus sistemas. Además, conocen de aquella información de uso recurrente entre las dependencias que definen algunos objetos transversales a la Universidad, por ejemplo, sedes, facultades, programas curriculares, personas, entre otros.

Adicionalmente, y dada la naturaleza de la Universidad, se requiere conocer los marcos normativos y los macroprocesos de la institución que definen las operaciones y transacciones de los sistemas de la información ya que esto permite dar un contexto y una justificación en términos de roles y responsabilidades, tiempos y ocurrencia de las transacciones, lo cual permite determinar si se gestionan datos maestros o no.

6.3.2. Entradas

El punto de referencia para la gestión de datos maestros y referencia es su definición e identificación, por ello es necesario establecer un contexto adecuado para la Universidad conociendo a profundidad su operación. Para lograr esto es fundamental analizar los siguientes insumos:

- Documentación de los procesos de la Universidad.
- Experiencia de los equipos técnico y funcional de los sistemas de la información corporativos, de acuerdo con el alcance establecido para la fase I.

Con el fin de identificar requerimientos para fortalecer la operación y gestión de la Universidad, además de enfocar una estructura que obedezca a un análisis de arquitectura de datos, es importante contar con los siguientes insumos:

- Diagnósticos de problemas identificados por debilidades en la definición de los datos que se gestionan en la Universidad.
- Alcance de la dimensión de arquitectura de datos sobre la forma en que deben estructurarse los datos de acuerdo con la dinámica de funcionamiento propia de la Universidad.

6.3.3. Actividades

En las siguientes cuatro secciones se detallan las actividades para el desarrollo de la dimensión de gestión de datos maestros y de referencia, estableciendo una metodología que promueve su uso para dar respuesta a las necesidades identificadas en la Universidad de acuerdo con el alcance establecido.

6.3.3.1. Implementar la metodología de identificación de datos maestros y de referencia

6.3.3.1.1. Identificar datos maestros y de referencia

Con el fin de identificar aquellos datos que puedan considerarse maestros y el análisis para establecer la necesidad de documentarlos como datos de referencia o no, es necesario analizar los siguientes ítems:

Tabla 21 Ítems para identificar datos maestros y de referencia

Pregunta	Dato maestro	Dato de referencia
¿La existencia del dato condiciona la generación de transacciones propias a los fines de cada macroproceso?	✓	✓
¿Los datos se modifican por transacciones diferentes a las de registro y actualización?	✗	✗
¿Los datos fluyen entre más de un Macroproceso?	✓	✓
¿El dominio de los datos es finito?	✗	✓

De acuerdo con las respuestas a estos ítems se pueden identificar los datos como maestros o de referencia de la siguiente manera: si la respuesta a las preguntas del primer y el tercer ítem es afirmativa y la respuesta al segundo ítem es negativa, los datos son candidatos para considerarse datos maestros. Se considerarán datos de referencia dependiendo del tamaño del conjunto, ya que el hecho de que sean finitos o infinitos los condiciona como tal. Las listas finitas son candidatas a estandarización en sus metadatos de valor y formato por medio de tablas de referencia

6.3.3.1.2. Establecer datos maestros y catalogar datos de referencia

Una vez consolidada una definición para datos maestros, se parte de la metodología propuesta en la dimensión de modelado y diseño de datos y se identifican unos ítems adicionales para la clasificación, dando como resultado la siguiente metodología:

- Se realiza un análisis normativo (modelado y diseño de datos).
- Se realiza una entrevista a los profesionales funcionales (modelado y diseño de datos).
- Se modelan las transacciones que se llevan a cabo en el sistema de la información (modelado y diseño de datos).
- Se identifican los puntos de inicio definidos como un tipo de relación, que registra o actualiza los datos de una entidad.

Se deben analizar los datos o entidades identificados con la metodología anteriormente mencionada y verificar si cumplen con los ítems mencionados en la sección 6.3.3.1.1. Ahora bien, dadas las repuestas a las preguntas de verificación, se podrá definir estas entidades y todos sus atributos como entidades y datos maestros respectivamente. Cuando estos últimos se identifican y clasifican, se procede a documentarlos en los instrumentos creados con esa finalidad y en los modelos de datos.

Dependiendo de la respuesta respecto al tamaño del conjunto de datos, se identifican candidatos a datos de referencia y se procede a verificarlos y catalogarlos mediante las preguntas de la **Tabla 22**.

Tabla 22 Ítems para clasificar datos de referencia

Pregunta	Únicamente Dato maestro	Dato de referencia documentado en lista de referencia simple	Datos de referencia documentado en lista de referencia cruzada
¿El proceso de creación o modificación de los datos se realiza de forma poco frecuente?	✗	✓	✓
¿Es necesario incluir a los datos en listas cerradas?	✗	✓	✓
¿La forma más simple de los datos es un código acompañado de una descripción en lenguaje natural?	✗	✓	✗
¿Existen diferentes codificaciones para el mismo concepto en otros sistemas de información?	✗	✗	✓

6.3.3.2. Establecer instrumentos de documentación de datos maestros y de referencia

6.3.3.2.1. Datos maestros

Como se mencionó en la identificación de datos maestros (sección 6.3.3.1.1), se parte de la metodología de la dimensión de modelado y diseño de datos, en la que se especifica cómo los datos maestros serán documentados y, en específico, se toman los campos de verificación de la sección 2 “Entidades”. Esto con el fin de clasificar de manera automatizada si es una entidad maestra o no. Para ello, en el formulario para el levantamiento de datos e información (**Anexo 5**) se presentan las siguientes preguntas:

- ¿Su existencia condiciona la operación del sistema de información?
- ¿Los datos se modifican por transacciones diferentes a las de registro y actualización?
- ¿Los datos fluyen entre más de un sistema de información?
- ¿El dominio de los datos es finito?

6.3.3.2.2. Datos de referencia

Una vez identificados los datos de referencia se creó un instrumento para documentarlos (**Anexo 9 - Formato para lista de referencia simple**) el cual debe tener como mínimo los siguientes campos (en caso de ser documentado en una tabla de referencia simple):

- Código: un código numérico o alfanumérico que identifica inequívocamente a un elemento.
- Nombre: identificación en lenguaje natural del código.
- Descripción: significado del dato de referencia para la Universidad.
- Normativa: normativa que rige el significado y/o el caso de uso para el dato de referencia.

Para documentar los datos de referencia en una tabla de referencia cruzada (**Anexo 10 - Formato lista de referencia cruzada**) son necesarios los siguientes atributos:

- Código asignado por el gobierno de datos: un código numérico o alfanumérico que identifica inequívocamente a un elemento, asignado por aquel sistema de la información o entidad que tiene el gobierno sobre el dato.
- Nombre del sistema de la información que tiene el gobierno del dato.
- Nombre: identificación en lenguaje natural del código.
- Descripción: significado del dato de referencia para la Universidad.
- Normativa: normativa que rige el significado y/o el caso de uso para el dato referencia.
- Columnas con la codificación dada para el mismo término por otros sistemas de información.

El resultado de documentar una tabla de referencia cruzada será una tabla que servirá de traductora entre sistemas de la información para el mismo concepto definido con diferentes códigos. Esto tendrá aplicaciones en cuanto a la interoperabilidad, publicación de datos, entre otros. Las tablas creadas deberán ser incluidas en el modelo de datos y repositorios de metadatos.

6.3.3.3. Incluir los datos maestros y de referencia en el modelo de datos y el repositorio de metadatos

Sin la existencia de los datos maestros no es posible generar las transacciones propias a los fines de cada macroproceso, por ello es necesario que ellos sean los primeros en estar correctamente documentados en aquellos instrumentos que tienen como fin mostrar de manera holística los datos dentro del sistema, ya sean modelos de datos que representan el sistema de información, glosarios, repositorios de metadatos, entre otros. Así mismo, al ser una subclasificación de los datos maestros, los datos de referencia tienen una importancia relevante en la Universidad y por ello también deberán ser incluidos en los mismos instrumentos donde se documentaron los datos maestros.

En el caso de los datos maestros se documentaron las entidades siguiendo lo propuesto en la metodología de modelado y diseño de datos, priorizando aquellos que en el formulario de recolección se clasificaron automáticamente como datos maestros luego de responder las preguntas de la sección 4.3.1 de este capítulo. Los datos de referencia se documentaron usando las tablas de referencia creadas en el instrumento (**Anexos 9 y 10**), estas tablas deberán ser incluidas en los modelos de datos.

Es necesario tener especial precaución al crear relaciones entre las tablas de referencia y otras tablas cuando sean documentadas en los modelos de datos, ya que es posible que se presenten múltiples

relaciones dada su trascendencia, lo que implica un riesgo al perder la claridad del modelo. Se recomienda modelarlas como islas y resaltar de manera especial los atributos llave con el fin de identificar sus relaciones bajo una rápida inspección visual.

La inclusión de datos de referencia en los repositorios de metadatos apoya la gestión y almacenamiento de datos estandarizados y a la vez es un apoyo para la integración de datos debido a que documenta las traducciones de diferentes codificaciones entre sistemas de la información para un mismo concepto. Los datos identificados como datos maestros condicionan la generación de las transacciones propias a los fines de cada macroproceso, por ello son los datos de mayor prioridad para propiciar su intercambio lo que implica su inclusión en los repositorios de metadatos.

6.3.4. Salidas

De acuerdo con el alcance establecido, los artefactos producidos en esta fase son los que se proponen para recoger y documentar los datos maestros y de referencia. Adicionalmente se cuenta con un documento guía para diligenciar cada una de las secciones que componen el artefacto y, para finalizar, un documento que muestra el propósito de cada una de estas secciones. Para ello se crearon los anexos de este capítulo:

- Anexo 9: Formato para lista de referencia simple
- Anexo 9A: Instructivo de diligenciamiento del formulario para la documentación de tablas de referencia simple
- Anexo 9B: Propósito de las secciones del formulario para la documentación de tablas de referencia simple
- Anexo 10: Formato para lista de referencia cruzada
- Anexo 10A: Instructivo de diligenciamiento del formulario para la documentación de tablas de referencia simple cruzada
- Anexo 10B: Propósito de las secciones del formulario para la documentación de datos de referencia cruzada

6.3.5. Usuarios

Los usuarios de los datos maestros y de referencia se relacionan con aquellos que deben tomar decisiones de priorización de la gestión de datos o donde la estandarización de datos impacta en los resultados, por ejemplo:

- Desarrollar nuevas aplicaciones, dado que deben definirse cuáles son los datos mínimos requeridos para la operación del sistema y en caso de realizar integraciones, establecer su estandarización.
- Administrar bases de datos, dado que deben priorizar la gestión de los datos maestros.
- Gestionar productos y servicios de información, dado que se deben establecer de forma adecuada la estandarización de algunos datos para presentar resultados que puedan integrarse a otras fuentes.

En general, los equipos técnicos y funcionales son usuarios de los datos maestros y de referencia debido a que dentro de sus actividades están las de priorizar y estandarizar datos permitiendo hacer análisis e identificar oportunidades de mejora.

7. GESTIÓN DE METADATOS

7.1. Conceptualización

La definición más difundida para metadatos es “datos de los datos” (DAMA Internacional, 2017), lo que demuestra la deficiencia y ambigüedad para entender la estructura de los datos, su propósito, su importancia, su estandarización y su gestión. La naturaleza estructurada de los metadatos es fundamental porque establece un orden, por eso es posible catalogar información. Los metadatos cumplen diferentes propósitos: documentación, búsqueda, autenticación, auditoría, evaluación de calidad, entre otros. La importancia de los metadatos radica en que administran el conocimiento de una organización. Ahora bien, ese conocimiento puede ser tan diverso y heterogéneo que estandarizar impone desafíos adicionales para que los datos sean más susceptibles de intercambio interno y externo. La desconexión de cualquiera de los elementos anteriores dificulta la posibilidad de gestionar los datos como un activo.

Los metadatos desempeñan un papel crucial que gira en torno a los datos. En cierto sentido, pueden ser entendidos como capas de descripción, de contexto y de conceptualización que recubren a los datos (núcleo). Sin estas capas no sería posible entender un valor como “4.5”, que podría hacer referencia a una temperatura, el promedio académico de un estudiante o incluso la valoración de una película. Siguiendo este ejemplo, la manera más sencilla de “recubrir” este dato sería con el uso de una etiqueta; por ejemplo, “promedio académico”. Sin embargo, dicha etiqueta carece de los recursos necesarios para acercarnos a la realidad del dato porque no proporciona información sobre cómo se calcula dicho valor, si es el promedio por periodo o es el promedio general de carrera, cuáles son los valores máximos y mínimos, cuándo fue actualizado por última vez o si es un dato público, privado o sensible. Considerando lo anterior, no puede existir una única “capa” de metadatos en torno a los datos, se articulan según su tipología y cada una cumple un propósito definido.

Los metadatos son esenciales para la gestión de los datos, así como para el uso de los datos. Las organizaciones producen y utilizan una gran cantidad de datos esta se debe documentar o la organización corre el riesgo de perder valiosos conocimientos de sí misma. Información descriptiva de los datos Revelan patrones, relaciones, y comportamientos Los datos que registran la información de los datos físicos Los procesos técnicos que usan los datos del negocio que se beneficia en uso de los datos Las reglas de negocio y políticas que se establecen alrededores de datos Las restricciones y estructuras lógicas con que se construyen una los datos que se almacenan La ubicación física de los datos La definición más común de Metadatos, sería "Datos de los datos".

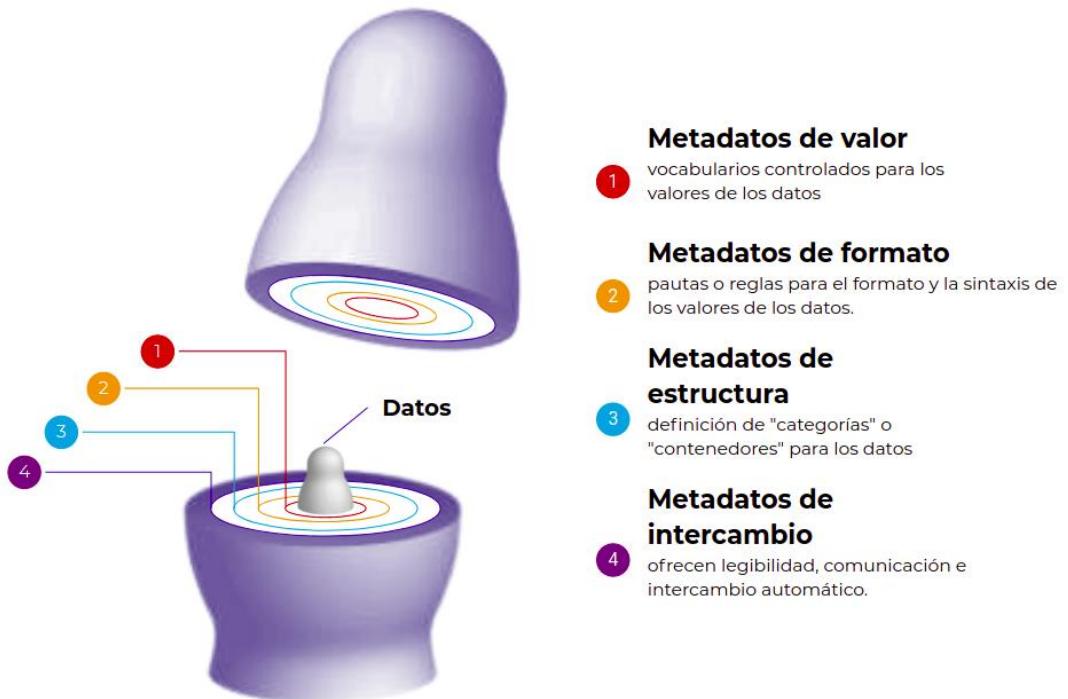
Los metadatos incluyen Para entender el papel vital de los Metadatos en la gestión de datos, imagínese una gran biblioteca, con cientos de miles de libros y revistas, pero sin catálogo de tarjetas. Sin un catálogo de tarjetas, los lectores no podrían ni siquiera saber cómo empezar a buscar un libro específico o incluso un tema específico. Los catálogos de tarjetas no solo proporcionan la información necesaria (Cuáles son los libros y materiales que la biblioteca posee y dónde está archivados), sino que también

permite a los clientes encontrar materiales utilizando diferentes puntos de partida (área temática, autor o título). Sin el catálogo, encontrar un libro específico sería difícil, si no imposible. Una organización sin Metadatos es como una biblioteca sin un catálogo de tarjetas.

7.1.1. Tipologías de metadatos

Las capas de metadatos que recubren un dato enriquecen su valor. La profundidad y alcance de cada capa está descrita en función de su propósito. Desde una perspectiva general existen cuatro capas estándar bien definidas en las que se pueden clasificar los metadatos (Gilliland, 2008). En la Figura 23 se ilustra esta clasificación usando una analogía con los cascarones que componen a una muñeca rusa (Sansone, 2017), también conocida como *matrioshka* o *bábushka*. De la revisión de la literatura se conforma una combinación de conceptos en el que las tipologías de metadatos se describen como cascarones (por lo menos cuatro) de la muñeca y los datos son el núcleo (la muñeca más interna).

Figura 23 Tipologías por capa para metadatos



- **Metadatos de valor:** proporcionan términos preestablecidos mediante vocabularios controlados para llenar un campo de datos (Gilliland, 2008). Por ejemplo, el valor para llenar el campo departamento_colombia no puede ser Cundinamarca_1, cundinamarca_\$, CUNDNMRC, Cundinamarca_Dpto. Este debe ser parte de un listado finito y estandarizado para referirse a la división geopolítica de Colombia como la codificación Divipola generada por

el DANE (**Figura 24**) para los valores que pueden tomar los nombres de los departamentos de Colombia, por lo cual el valor establecido correspondería a CUNDINAMARCA.

Figura 24 Ejemplo de identificación de metadatos de valor y estructura para la codificación de departamentos de Colombia DIVIPOLA



El futuro es de todos **gobierno de Colombia**

Codificación de la División Político Administrativa de

Departamentos

Código	Nombre
05	ANTIOQUIA
08	ATLANTICO
11	BOGOTÁ, D.C.
13	BOLÍVAR
15	BOYACÁ
17	CALDAS
18	CAQUETÁ
19	CAUCA
20	CESAR
23	CORDOBA
25	CUNDINAMARCA
27	CHOCÓ
41	HIUILLA
44	LA GUAJIRA
47	MAGDALENA
50	META
52	NARIÑO
54	NORTE DE SANTANDER
63	QUINDÍO
66	RISARALDA
68	SANTANDER
70	SUCRE
73	TOLIMA
76	VALLE DEL CAUCA
81	ARAUCA
85	CASANARE
86	PUTUMAYO
88	ARCHIPIÉLAGO DE SAN ANDRÉS, PROVIDENCIA Y SANTA CATALINA
91	AMAZONAS
94	GUAINÍA
95	GUAVIARE
97	VAUPES
99	VICHADA

Fuente: DANE.
Nota: En la actualidad el país se compone de 32 departamentos y adicionalmente para fines estadísticos se codifica como departamento el Distrito Capital de Bogotá.
Actualizado al: 26 de Marzo de 2021

- **Metadatos de formato:** codificación particular de pautas o reglas para el formato y la sintaxis de los valores de los datos (Gilliland, 2008). Siguiendo el ejemplo anterior, en la actualidad el país se compone de 32 departamentos y adicionalmente para fines estadísticos se codifica como departamento el Distrito Capital de Bogotá. Cada departamento tiene asociada una codificación única y una lista finita de valores para otros atributos (**Figura 25**); por ejemplo, el nombre tal como se presentó en el ejemplo de metadatos de valor. Para establecer el mejor formato que contribuya a su almacenamiento y captura es necesario definir el formato de dos campos, uno para las codificaciones y otro para los nombres. En este caso, puede analizarse el contenido de la codificación Divipola e inferir el formato de cada uno de los campos y posteriormente deben buscarse elementos en otras capas de metadatos que aporten a la definición del formato.

Figura 25 Ejemplo de identificación de metadatos de formato y estructura para la codificación de departamentos de Colombia DIVIPOLA



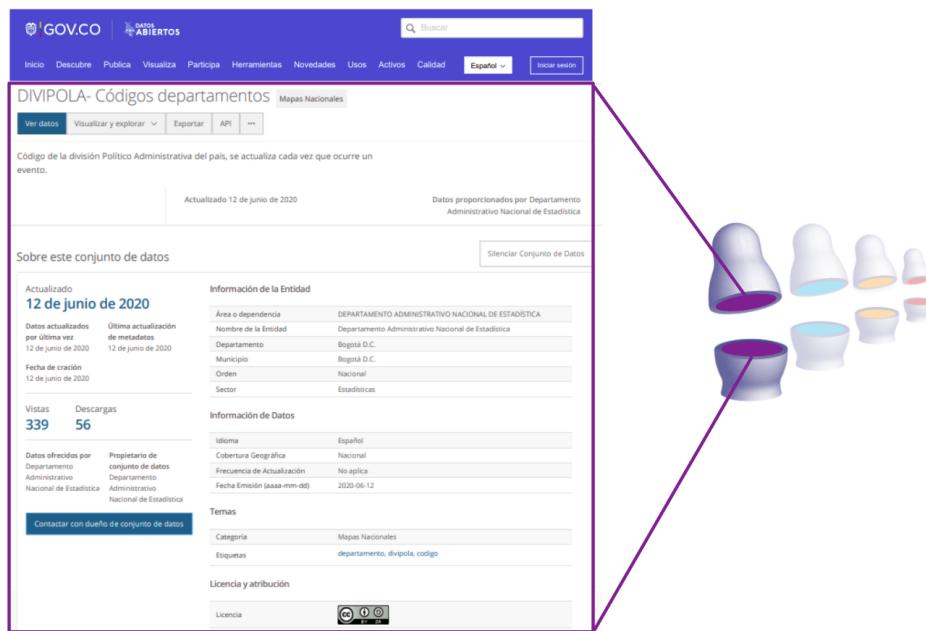
- **Metadatos de estructura:** definen los contenedores para los datos; es decir, aquellos conjuntos de elementos que deben ser identificados y definidos bajo un orden para facilitar la captura de los datos (Gilliland, 2008). Hasta este punto, los metadatos de formato y valor no son suficientes para describir el contexto total del dato. En el ejemplo que estamos trabajando, sabemos que el dato hace referencia al nombre de un departamento, pero no es claro si corresponde al departamento donde está ubicada una sede, el departamento de residencia o el departamento de procedencia de una persona.

En esta capa se utilizan etiquetas y atributos adicionales mínimos que debe llevar el dato para enriquecerlo; por ejemplo, “Departamento de procedencia”, y se agregan atributos como: descripción, fecha de actualización, creador, abreviación, entre otros. Los metadatos de estructura pueden asociarse con las etiquetas o textos de ayuda cuando se crea un formulario o al nombre y comentarios de ayuda cuando se definen los encabezados en una tabla de Excel.

- **Metadatos de intercambio:** ofrecen legibilidad, comunicación e intercambio automático (Gilliland, 2008). Los metadatos de intercambio hacen referencia al protocolo particular utilizado para compartir datos. De acuerdo con el ejemplo, si el propósito es publicar un catálogo de datos de los departamentos de procedencia de las personas asociadas a la Universidad en el portal de datos abiertos colombiano⁶ (**Figura 26**), se deben proporcionar los metadatos mínimos requeridos para que se pueda reutilizar el contenido de acuerdo con el vocabulario DCAT que actualmente es el requerido por dicho portal y el definido por el Ministerio de las Tecnologías de la Información y las Comunicaciones (2019) como parte de la estrategia de Gobierno en Línea (2015).

⁶ <https://www.datos.gov.co/>

Figura 26 Ejemplo de identificación de metadatos de intercambio para la codificación de departamentos de Colombia DIVIPOLA

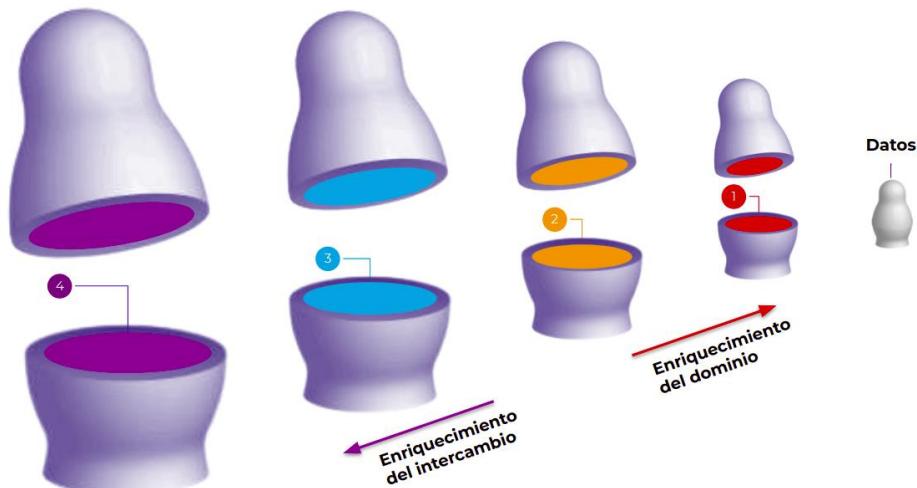


The screenshot displays the DIVIPOLA Data Catalogue interface. Key visible information includes:

- Actualizado:** 12 de junio de 2020
- Información de la Entidad:**
 - Área o dependencia: DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA
 - Nombre de la Entidad: Departamento Administrativo Nacional de Estadística
 - Departamento: Bogotá D.C.
 - Municipio: Bogotá D.C.
 - Orden: Nacional
 - Sector: Estadísticas
- Información de Datos:**
 - Idioma: Español
 - Cobertura Geográfica: Nacional
 - Frecuencia de Actualización: No aplica
 - Fecha Emisión (aaaa-mm-dd): 2020-06-12
- Temas:**
 - Categoría: Mapas Nacionales
 - Etiquetas: departamento, divipola, codigo
- Licencia y atribución:**
 - Licencia: CC BY SA

En la **Figura 27** se puede observar una vista desglosada donde la parte más interna, es decir, el núcleo, está compuesto por los datos. Por una parte, entre más interna sea una capa, más se enriquece el dominio de los datos (valor y formato). Por otra, entre más externa sea una capa más se fortalece la posibilidad de uso e intercambio de datos con criterios de calidad. Todas las capas son importantes y en cada una de ellas se busca garantizar la calidad de los datos en su interior y finalmente favorecer la comunicación entre personas y máquinas.

Figura 27 Vista detallada de cada una de las capas de metadatos



7.1.2. Categorías de metadatos

Un término que resuena bastante en el ámbito de los metadatos es el contexto. El contexto indica los aspectos de quién, qué, dónde, por qué, cuándo y cómo, asociados a los datos. Entre mejor sea el contexto, menor es la probabilidad de que se hagan suposiciones sobre los datos. Para entender mejor esta visión se necesita hacer una distinción de los metadatos sin importar la capa en la que nos encontramos con respecto al dato y es por lo que los metadatos se pueden categorizar según su función.

En la **Tabla 23** se consignan preguntas típicas de contextualización que dan origen a tres categorías: metadatos de negocio, técnicos y operacionales, y de gobierno. En general, su categorización se da en función de la afinidad de las preguntas; por ejemplo, es evidente que los metadatos de gobierno tienen en común al quién.

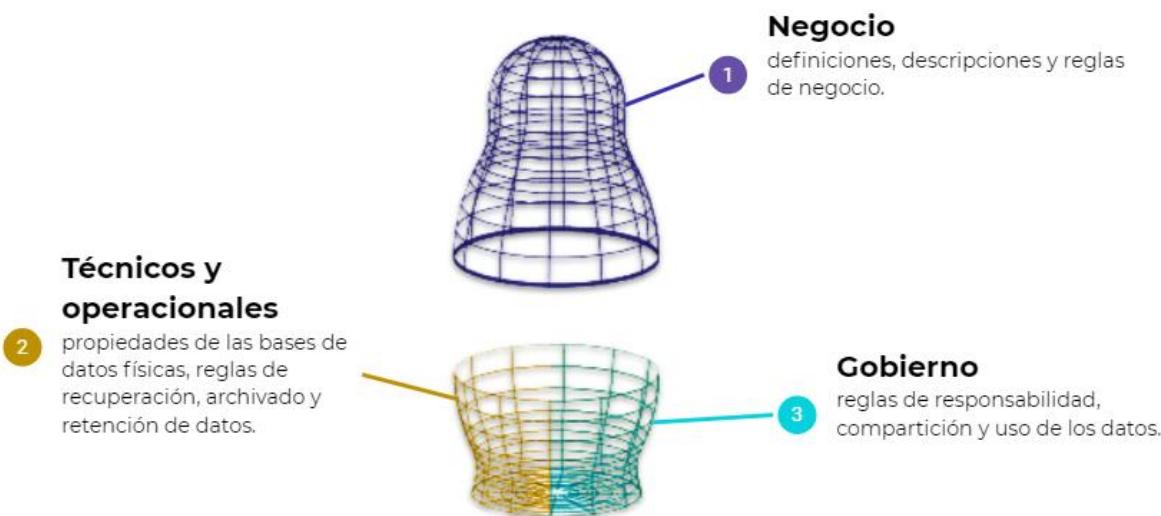
Tabla 23 Preguntas de contextualización para metadatos (Burbank, 2020)

Quién	Qué/Cuáles	Dónde	Por qué	Cuándo	Cómo
¿Quién ha creado estos datos?	¿Qué definición empresarial tiene este elemento de datos?	¿Dónde se almacenan estos datos?	¿Por qué se almacenan estos datos?	¿Cuándo se crearon estos datos?	¿Cómo están codificados estos datos? (carácter, numérico, etc.)
¿Quién es el responsable de estos datos?	¿Cuáles son las reglas de negocio para estos datos?	¿De dónde proceden estos datos?	¿Por qué se utilizan y para qué sirven?	¿Cuándo se actualizaron estos datos por última vez?	¿Cómo se almacenan estos datos en las bases de datos o en las fuentes de datos?
¿Quién utiliza estos datos?	¿Qué nivel de seguridad o privacidad tienen estos datos?	¿Dónde se utilizan y comparten estos datos?		¿Durante cuánto tiempo deben almacenarse?	¿Cómo se restringen los valores válidos?
¿Quién es el dueño de estos datos?	¿Qué abreviatura o acrónimo tiene este elemento de datos?	¿Dónde está la copia de seguridad de estos datos?		¿Cuándo hay que purgar/eliminar?	
¿Quién regula o añade estos datos?	¿Qué normas técnicas existen para la implementación de la base de datos?	¿Existen políticas regionales de privacidad o seguridad que regulen estos datos?		¿Cuándo se generan nuevos reportes?	

En la **Figura 28** se ilustran las categorías de metadatos enmarcadas en las tipologías, esto es, las capas pueden estar compuestas de metadatos de negocio, técnicos y operacionales, y de gobierno. Es importante mencionar que las proporciones de esa composición pueden ser diferentes en cada capa; por ejemplo, las capas de valor y formato pueden tener mayor proporción de metadatos técnicos y operacionales, en cambio, en las capas de estructura e intercambio predominarán metadatos de negocio.

y de gobierno. En conclusión, las categorías de metadatos nos permiten definir el origen de los metadatos y las tipologías nos ayudan a establecer su aprovechamiento.

Figura 28 Representación de la categorización de metadatos



- **Metadatos de negocio:** También conocidos como administrativos o descriptivos. Contienen la información de las definiciones, descripciones y reglas de negocio. Establecen niveles de seguridad y privacidad. Sirven para fijar reglas de calidad y horarios de actualización. Formalizan el origen y linaje de los datos.
- **Metadatos técnicos y operacionales:** Suministran información detallada de las propiedades de las bases de datos físicas, permisos de acceso, registro de errores, reglas de respaldo, recuperación, archivado y retención de datos.
- **Metadatos de gobierno:** Proporcionan información de contacto de las partes interesadas (por ejemplo, propietarios de datos, administradores de datos). Especifican reglas de responsabilidad, compartición y uso de los datos.

La implementación de categorías y tipologías, así como la definición y aplicación de estándares concertados y apropiados por todos los procesos de la Universidad, permitirá la recuperación de la información de manera eficiente.

7.2. Propósito

La categorización, documentación y acopio de metadatos sobre los componentes de la información (sean documentos, bases de datos, definiciones, y cualquier otro tipo de datos que se produzcan en la Universidad) fortalecen la gestión de conocimiento en la institución y permiten determinar el inventario de activos de datos e información existente. Gestionar metadatos encierra las actividades de

planificación, implementación y control para permitir el acceso y la administración de los metadatos que describen los componentes de la información de la entidad (DAMA Internacional, 2017).

7.2.1. Metas

- Documentar y gestionar el conocimiento propio de la Universidad, en lo que al entendimiento de la institución se refiere, con la terminología que se sustenta en sus marcos normativos. En consecuencia, la comunidad universitaria puede apropiar el contenido de los datos y utilizarlos de forma consistente.
- Determinar los estándares adecuados para permitir el acceso e intercambio de metadatos, analizando las necesidades de acuerdo con los componentes de datos e información que se establezcan.

7.2.2. Alcance fase II

La diversidad de componentes de datos e información y, por ende, de necesidades de metadatos para clasificar y describir dichos componentes, es bastante elevada en la Universidad. Esto implica un cambio de paradigmas en la institución, instaurando que la cultura de documentación sea vista de forma complementaria a la producción y mantenimiento de datos necesarios para la operación cotidiana, donde lo que se conoce de los datos (su definición, uso, casos especiales, entre otros) deba ser documentado para que dicho conocimiento (la suma de los datos y los metadatos) pueda ser transferido.

En el desarrollo de esta dimensión se establecerá el punto de partida conceptual y metodológico para identificar y clasificar metadatos que puedan aportar a la mejora de la gestión de la información de acuerdo con sus tipologías y categorías. Como parte de la estrategia de metadatos se propone que se establezca si se adoptarán o no estándares específicos en cada una de las tipologías definidas ya que debe ser un proceso concertado y reconciliado con los sistemas de la información existentes.

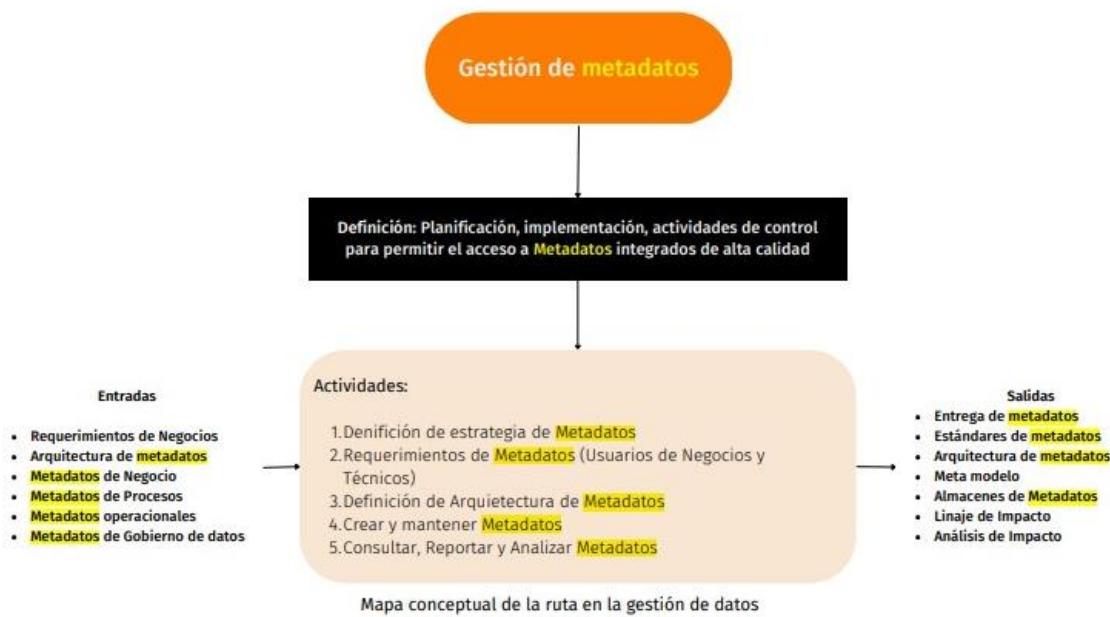
7.3. Desarrollo de la fase II de gestión de metadatos

Tabla 24 SIPOC para la dimensión de gestión de metadatos en la fase II

Proveedores	Entradas	Actividades	Salidas	Usuarios
Equipos técnicos de sistemas de información	Diagnósticos de sistemas de información, marco normativo de procesos	Entender requerimientos de metadatos	Identificación de requerimientos de metadatos	DNED
DNED	Identificación de requerimientos de metadatos	Definir estrategia de metadatos	Estrategia de metadatos	Comité de Gobierno de la Información, Gestores de productos y servicios de información

DNED	Definir estrategia de metadatos	Proponer instrumentos de documentación de metadatos	Propuesta de instrumentos de documentación en metadatos	Equipo técnico y funcional de los sistemas de información, Analistas de calidad de datos
------	---------------------------------	---	---	--

Figura 29 Mapa conceptual de la gestión de la ruta en la gestión de metadatos



Los estándares de metadatos son documentos formales que establecen criterios, métodos, procesos y prácticas uniformes, han sido creados con el fin de guiar el diseño, la creación e implementación de estructuras de datos, valores de datos, contenidos de datos e intercambio de datos de manera eficiente y consistente. Dependiendo del propósito para el cual el estándar ha sido desarrollado puede dividirse en cuatro categorías en el contexto de las LAM (librerías, archivos y museos.):

Estándares para estructuras de datos: Usualmente conocidos como "conjuntos de elementos" o "vocabularios de metadatos" estos estándares definen estructuras de datos y semántica. Entre ellos se encuentran: Dublin Core VRA Core EAD MARC21 Estándares para contenido de datos: Los estándares de contenido de datos son creados para guiar las prácticas de generación de metadatos y catalogación. Algunos de ellos son: RDA AACR2 CCO Estándares para valores de datos: Hay diferentes clases de sistemas de organización del conocimiento, generalmente conocidos como "vocabularios de valores" (value vocabularies) y a veces como "esquemas de codificación de valores" (value encoding schemes). Algunos de ellos son: LCSH AAT TGN DDC Estándares para intercambio de datos: Estos son conocidos como "formatos" cuando se discuten en el contexto del intercambio de datos y la comunicación. Algunos ejemplos son: ISO2709 (MARC) XML JSON html5

7.3.1. Proveedores

Para el desarrollo de la dimensión se requiere, por una parte, un proceso de actualización permanente sobre definiciones y gestión de metadatos, teniendo en cuenta que la gestión de la información es una rama del conocimiento en permanente evolución. Por otra parte, se requiere integrar esfuerzos con los profesionales miembros de las dependencias que tienen bajo su responsabilidad la parte técnica y funcional de los sistemas de la información corporativos de la Universidad, o que actúan como interlocutores funcionales para el soporte técnico de las soluciones que se tienen tercerizadas a un proveedor ya que ellos conocen el contexto, el uso y aplicación de los datos en la Universidad.

Adicionalmente y dada la naturaleza de la Universidad, se requiere conocer los marcos normativos y los macroprocesos que definen las operaciones y transacciones de los sistemas de información, ya que esto permite identificar los mínimos elementos de metadatos existentes.

7.3.2. Entradas

El punto de referencia para la gestión de metadatos es su definición e identificación, por ello es necesario establecer un contexto adecuado para la Universidad conociendo a profundidad su operación. Para lograr esto es fundamental analizar los siguientes insumos:

- Documentación de los procesos de la Universidad.
- Experiencia de los equipos técnico y funcional de los sistemas de la información corporativos de acuerdo con el alcance establecido para la fase I, con el fin de reunir metadatos de negocio, técnicos, operacionales y de gobierno.
- Las tablas de referencia simple y las tablas de referencia cruzada generadas en la dimensión de maestros y referencia con el fin de identificar elementos susceptibles de estandarización.

7.3.3. Actividades

A lo largo de la gestión de datos es usual preguntarse qué datos tiene la Universidad, qué representan, cómo se clasifican, de dónde proceden, cómo se mueven dentro de la Universidad, cómo evolucionan a través de su uso, quién puede y quién no puede usarlos y cómo se puede evaluar su calidad (DAMA Internacional, 2017). Las actividades presentadas en esta dimensión son una forma de dar respuesta a las anteriores preguntas. Estas ofrecen un punto de partida para contextualizar y entender los conjuntos de datos de los sistemas de información. Precisamente, la contextualización comienza con las definiciones y descripciones consignadas en lo que se conoce como glosario de negocio. Sin embargo, un glosario por sí solo no es una demostración de que se está haciendo gestión y existen otras aristas que suelen surgir del trabajo con metadatos. Por lo anterior, en esta sección se especifican las actividades para estructurar acciones entorno a la gestión de metadatos.

7.3.3.1. Entender los requerimientos de metadatos

A corto plazo, la gestión de metadatos permite evitar la duplicación de datos, compartir información fiable; buscar, recuperar y evaluar la información de los conjuntos de datos dentro y fuera de la Universidad; a largo plazo, su reutilización ahorra tiempo y recursos (DataONE, 2016). Los metadatos no solo existen para encontrar datos, también contribuyen a decidir si un conjunto de datos satisface una necesidad concreta, ayudan a descubrir, procesar y utilizar el conjunto de datos y aceleran el proceso de compresión (por ejemplo, definiciones de los nombres de los atributos, o dominios de los datos). La materialización de los beneficios a corto y largo plazo se hace con la ayuda de los requerimientos de metadatos.

Para empezar a entender los requerimientos de metadatos se reconoce que no todos tienen el mismo efecto sobre la conexión entre los conjuntos de datos y las reglas de negocio, sus definiciones y las formas de representarlos. Esto significa que el esfuerzo y la energía se centrará en la gestión de aquellos que tienen mayor repercusión en la Universidad y sus sistemas de información. Sin embargo, definir la magnitud que tendrá la definición de un requerimiento posee demasiadas implicaciones en esta etapa, es por eso por lo que los primeros pasos para una valoración cualitativa se dan sobre una serie de preguntas de reflexión:

- ¿Para qué conjunto de datos quiere gestionar metadatos ahora?
- ¿Para qué conjunto de datos quiere gestionar metadatos en el futuro?
- ¿Para qué conjunto de datos ha identificado la imposibilidad de gestionar metadatos?
- ¿Cuáles son sus casos de uso actuales para los metadatos?
- ¿Cuáles son sus casos de uso futuros para los metadatos?
- ¿Cuáles son sus casos de uso para mostrar las relaciones entre las tipologías y categorías de metadatos?
- ¿El conjunto de datos se va a compartir o es solamente para su sistema de información?
- ¿Desea responder a los requerimientos de metadatos a entidades oficiales externas (SNIES, MinEducación, Minsalud)?
- ¿Debe hacer transformaciones al conjunto de datos para compartir información con entidades externas?
- ¿Alguna vez ha encontrado inconsistencias en el conjunto de datos utilizado?

Las respuestas a estas y otras preguntas irán definiendo un camino para reconocer el estado actual y el futuro para la gestión de metadatos basado en requerimientos que permitan conectar cualquier conjunto de datos con posibles usos y aplicaciones de una manera organizada y eficiente. Dicha conexión se da en términos de los principios de finalidad, restricciones de uso, mantenimiento, calidad, integración y gobierno que explicamos a continuación.

7.3.3.1.1. Principios para requerimientos de metadatos

Los principios son la base para reconocer la importancia de la gestión de metadatos. Enfatizan que los metadatos deben tener un propósito, no pueden revelar información de datos protegidos, pertenecen a

un ciclo en el que no solo son creados sino también actualizados, proporcionan confianza, pueden ser unificados y hacen parte de una cultura organizacional y de gestión de cambio que reconoce y garantiza los recursos necesarios para la gestión de información.

- ***Finalidad:***

Se refiere al propósito previsto por el dueño de los datos, ya que es quien mejor conoce el ciclo de estos (origen, actualización, etc.) y tiene la capacidad de documentar y gestionar el conocimiento de la terminología relacionada con el conjunto de datos para garantizar que los consumidores entiendan su contenido y puedan utilizarlos de forma coherente. Por ejemplo, si a un consumidor le entregan un conjunto de datos con el índice de resiliencia epidemiológica municipal (IREM), éste podría preguntarse a qué hace referencia este índice, cómo se calcula y cuáles son sus unidades. Si el dueño sigue el principio de finalidad, entonces entregará metadatos de descripción o definición porque tienen como propósito proporcionar una explicación en lenguaje natural de los datos. En la **Figura 30** se encuentra una porción de los metadatos entregados por MinSalud (2021) donde se encuentra la definición ampliada del IREM, la fórmula de cálculo e incluso las fuentes de información.

Figura 30 Ficha técnica sobre el índice de Resiliencia Epidemiológica Municipal - parte 1 (Ministerio de Salud, 2021)

Nombre del indicador	Índice de Resiliencia Epidemiológica Municipal (IREM)										
Definición	<p>Se trata de un índice sintético multidimensional conformado por tres dimensiones, a saber: (i) avance en las coberturas de vacunación contra el COVID-19 en la población a partir de los 16 años; (ii) estimación de la seroprevalencia del SARS-CoV-2 en el municipio, ajustada por la razón de juventud; y (iii) capacidad del sistema de salud en el territorio.</p> <p>El índice varía entre 0 y 1 puntos. Cuando el índice tiende a 1, se concluye una mayor resiliencia epidemiológica del municipio ante la apertura económica, cultural y social en el marco de la superación de la pandemia por COVID-19. Los detalles metodológicos para la construcción del índice se especifican en la ficha técnica.</p>										
Fórmula de cálculo	<p>IREM =</p> $0,5 \text{ (Índice vacunación)} + 0,3 \text{ (Seroprevalencia} \times R\text{)} + 0,15 \text{ (Capacidad Sistema Salud)} + 0,05 \text{ (Testeo promedio al mes)}$ <p>Índice de vacunación =</p> $\frac{(\% \text{ población} \geq 16 \text{ con una dosis} \times 0,4) + (\% \text{ población} \geq 16 \text{ con esquema completo} \times 0,6)}{}$ <p>Seroprevalencia =</p> $\frac{[\text{Número de muertes por COVID} - 19 \text{ en el mpio}]}{\text{Letalidad en los infectados en el mpio}} / \text{Población proyectada en el mpio}$ <p>Razón de Juventud (RJ) =</p> $\frac{\text{Proporción de población} < 40 \text{ años en el municipio}}{\text{Proporción de población} < 40 \text{ años en el país}}$ <p>Capacidad del Sistema de Salud en el territorio =</p> $\frac{\text{Número de camas instaladas UCI} + \text{Intermedio}}{\text{Población en el municipio}}$ <p>Testeo promedio al mes (considerando los últimos 3 meses) =</p> $\frac{\text{Número de pruebas (PCR} + \text{antígeno) realizadas en el mpio durante los 3 meses anteriores}}{3} / \text{Población del municipio}$										
Fuentes de información	<table border="1"> <tr> <td>Vacunas aplicadas</td> <td>Reporte diario de vacunación. Dirección de Promoción y Prevención. Ministerio de Salud y Protección Social</td> </tr> <tr> <td>Población</td> <td>Proyecciones y retroproyecciones de población municipal por área para 2021. DANE. Disponible en: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion</td> </tr> <tr> <td>Seroprevalencia</td> <td>Instituto Nacional de Salud (INS). Estimaciones actualizadas y reportadas quincenalmente</td> </tr> <tr> <td>Capacidad del sistema en el territorio</td> <td>Información suministrada por la Dirección de Prestación de Servicios. Ministerio de Salud y Protección Social</td> </tr> <tr> <td>Testeo</td> <td>Instituto Nacional de Salud (INS). Consulta en el cubo Sismuestras. SISPRO</td> </tr> </table>	Vacunas aplicadas	Reporte diario de vacunación. Dirección de Promoción y Prevención. Ministerio de Salud y Protección Social	Población	Proyecciones y retroproyecciones de población municipal por área para 2021. DANE. Disponible en: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion	Seroprevalencia	Instituto Nacional de Salud (INS). Estimaciones actualizadas y reportadas quincenalmente	Capacidad del sistema en el territorio	Información suministrada por la Dirección de Prestación de Servicios. Ministerio de Salud y Protección Social	Testeo	Instituto Nacional de Salud (INS). Consulta en el cubo Sismuestras. SISPRO
Vacunas aplicadas	Reporte diario de vacunación. Dirección de Promoción y Prevención. Ministerio de Salud y Protección Social										
Población	Proyecciones y retroproyecciones de población municipal por área para 2021. DANE. Disponible en: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/proyecciones-de-poblacion										
Seroprevalencia	Instituto Nacional de Salud (INS). Estimaciones actualizadas y reportadas quincenalmente										
Capacidad del sistema en el territorio	Información suministrada por la Dirección de Prestación de Servicios. Ministerio de Salud y Protección Social										
Testeo	Instituto Nacional de Salud (INS). Consulta en el cubo Sismuestras. SISPRO										

- Restricciones de uso:**

Está relacionado con la seguridad: algunos metadatos no deben ser expuestos porque revelarían la existencia de datos altamente protegidos. Por ejemplo, si un sistema de la información opera bajo acuerdos contractuales con proveedores externos, es natural que existan restricciones de uso respecto al modelo de datos y el motor del sistema que finalmente son propiedad intelectual del proveedor. Consideremos un segundo ejemplo: los datos que especifican las rutas de recolección de dinero de una empresa de transporte de valores deben ser altamente protegidos y sus metadatos no pueden ser expuestos; exponer un metadato tan simple como la fecha de actualización de las rutas podría poner en riesgo la operación. Finalmente, existen metadatos que le entregan al consumidor una idea sobre el uso que le puede dar a los datos a través de la especificación de un tipo de licencia, de marca registrada o vigencia de una patente.

- **Mantenimiento:**

Define los procesos y reglas para la actualización de metadatos. Regresando al ejemplo del IREM, es usual preguntarse como consumidor la periodicidad de actualización de los datos. MinSalud (2021) en este caso entrega un metadato que informa que la periodicidad de los datos es quincenal y la fecha más reciente de actualización, al momento de escribir este documento, es el primero de junio de 2021 y corresponde a la primera versión del documento (**Figura 31**).

Figura 31 Ficha técnica sobre el índice de Resiliencia Epidemiológica Municipal - parte 1 (Minsalud, 2021)

Unidad de medida	Puntos	Límites	0 a 1
Niveles de desagregación	Municipios capitales	Periodicidad	Quincenal
Responsable del cálculo	Dirección de Epidemiología y Demografía. Ministerio de Salud y Protección Social		
Notas técnicas	<ul style="list-style-type: none"> - El valor máximo que podrá tomar el producto entre la seroprevalencia y la razón de juventud será la unidad. - La capacidad del Sistema de Salud en el Territorio y el testeo promedio al mes se incluyen al índice después de ser estandarizados entre 0 y 1, utilizando la siguiente fórmula: $\text{índicador estandarizado} = \frac{\text{Valor observado}-\text{Valor mínimo posible}}{\text{Valor máximo posible}-\text{Valor mínimo posible}}$ <ul style="list-style-type: none"> - Para la capacidad del sistema el valor mínimo y máximo provienen de los valores observados (0,023 y 0,102, respectivamente), entre tanto, para el testeo, se tomará como valor mínimo el 0 y máximo el 10 - Este indicador contiene los parámetros básicos para la medición de la resiliencia epidemiológica municipal. De acuerdo con la evolución de la pandemia y con la disponibilidad de mayor evidencia científica relacionada con el virus y la enfermedad, el índice podrá ser actualizado y adaptado metodológicamente. 		
Versión: 1	Fecha	01-06-2021	Elaboró: mrojasb Dirección de Epidemiología y Demografía

- **Calidad:**

Proporciona un desarrollo confiable e inter organizativo. Permite una comprensión coherente de los valores de los datos. Por ejemplo, en situaciones donde hay ambigüedades respecto a los valores, los metadatos sirven para identificar inconsistencias y ayudan a encontrar valores no esperados, como sería el caso de encontrar direcciones de correos electrónicos en campos que almacenan números telefónicos. Como segundo ejemplo, en el caso de los datos IREM, es importante observar que el índice puede tomar solo valores numéricos entre 0 y 1, que corresponden a los límites definidos por el dueño.

- **Integración:**

Define el nivel de unificación de metadatos de diferentes fuentes y sus normas para entrelazarlos. Por ejemplo, cuando no se aplica este principio y existen múltiples consumidores, se producen diferentes resultados de análisis cuando ellos solicitan datos de una misma fuente, esto da lugar a malas interpretaciones debido a que los metadatos de descripción no están unificados. Usando el ejemplo del IREM, el cálculo del índice requiere la integración de diferentes fuentes: reporte diario de vacunación de

MinSalud, proyecciones del DANE (Departamento Administrativo Nacional de Estadística), estimaciones de seroprevalencia del INS (Instituto Nacional de Salud), capacidad del sistema en el territorio y el número de pruebas realizadas (ver metadatos de fuentes de la información en la **Figura 30**). Sin ese nivel de integración no sería posible entender la fórmula para hacer el cálculo del IREM que requiere específicamente de los datos de las fuentes mencionadas para llegar a obtener un valor.

- **Gobierno:**

Manifiesta de forma explícita la responsabilidad del dueño de los datos y/o responsabilidad asumida por el custodio. Por ejemplo, de los metadatos del IREM se puede identificar claramente que la fuente oficial de los datos es el Ministerio de Salud (2021), este es quien les da un propósito a los metadatos, establece las restricciones de uso, define reglas de mantenimiento, entrega información confiable y establece mecanismos de integración.

Estos principios no son los únicos, pero sirven de base para impulsar el enfoque que se debe tener para la definición de requerimientos (Queensland Goverment, 2019). Siguiendo esta línea, si bien estos principios se pueden aplicar a cualquier conjunto de datos se recomienda dar prioridad:

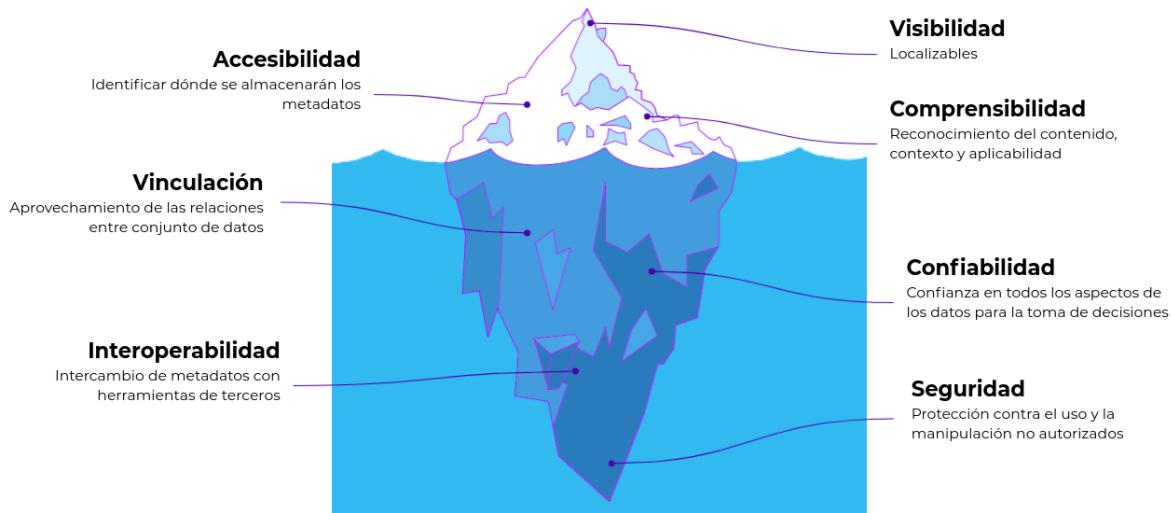
- Si son utilizados por más de un sistema de la información (por ejemplo, datos maestros).
- Si son reutilizables (por ejemplo, datos de referencia).
- Si responden a un requerimiento identificado previamente (por ejemplo, están estandarizados).
- Si se cuenta con herramientas tecnológicas para su gestión.
- Si son datos estructurados.
- Si apoyan la existencia de un macroproceso; es decir, corresponden a un activo sostenido y no está previsto que se retire su uso en el futuro.
- Si se utiliza para los productos o servicios de información. Si ya existe un producto, éste debería ir con metadatos.

Aun cuando se lleguen a aplicar los principios de finalidad, calidad, integración, etc., es preciso fijar metas para medir el avance en el cumplimiento de los requerimientos. Lo anterior obedece a que por más que se defina la finalidad de un metadato, si no es por lo menos comprensible, visible o accesible, cualquier persona que quiera hacer uso de él encontrará problemas para su aprovechamiento.

7.3.3.1.2. Metas de los requerimientos

Las metas de los requerimientos abogan para que los metadatos, sin importar su tipología o su categoría, sean localizables, comprensibles, estén conectados, puedan ser reutilizados, se garantice su disponibilidad, se reconozca su contenido, contexto y aplicabilidad (U.S. Department of Defense, 2020). Las primeras metas que saltan a la vista son las de visibilidad, accesibilidad y comprensibilidad, que, como en la punta de un iceberg, son aquellas a las que se les puede dar mayor atención. Sin embargo, existen otras metas cuya repercusión puede provocar el éxito o fracaso de la gestión y no siempre suelen ser consideradas a la hora de formular requerimientos de metadatos (**Figura 32**).

Figura 32 Representación de metas para requerimientos de metadatos



● **Visibilidad:**

Los consumidores pueden localizar los datos necesarios para un servicio o aplicación. Esta visibilidad se soporta porque existe un registro histórico del origen de los datos para identificar su procedencia y existen mecanismos para administrar su control de cambios. Los factores para evaluar la visibilidad son:

- Los metadatos son explícitos.
- Existe algo que indique que existen metadatos.
- Los metadatos están documentados.

● **Accesibilidad:**

Los metadatos deben ser fáciles de acceder tanto por personas (pueden asociar metadatos con sus respectivos datos) como por máquinas (automatización de la extracción y del intercambio de metadatos en diversos sistemas de información). Los consumidores pueden recuperar los datos porque identifican dónde se almacenan los metadatos. Se enriquece el uso de los datos mediante el descubrimiento automático y catalogado de conjuntos de datos con el uso de herramientas IA (Inteligencia Artificial) / ML (Machine Learning) para apoyar actividades más amplias de gestión de datos en el futuro. Los factores para evaluar la accesibilidad son:

- Es posible encontrar los metadatos.
- Existe un medio para hacer consulta de metadatos, bien sea automático o manual.
- Los metadatos incluyen clara y explícitamente identificadores de su tipología y/o categoría.
- Los metadatos se registran o indexan en un registro de metadatos.
- Los metadatos son recuperables por su identificador utilizando un protocolo de comunicación estandarizado.

● **Comprendibilidad:**

Los consumidores pueden encontrar descripciones de los datos para reconocer el contenido, contexto y aplicabilidad. Existe una semántica empresarial para apoyar las variaciones semánticas en los glosarios y la terminología de los conjuntos de datos. Un usuario, con o sin experiencia, puede comprender el conjunto de datos porque se tienen descripciones adecuadas. Los factores para evaluar la comprensibilidad son:

- Los metadatos se describen de forma rica con una pluralidad de atributos precisos y relevantes en cada una de sus categorías.
- Los metadatos están asociados a una procedencia detallada.
- Los metadatos cumplen con las normas de la Universidad.
- Los metadatos apoyan la comprensión e interpretación de los datos que describen.
- Los metadatos dan el contexto suficiente para entender y utilizar los datos que acompañan.
- Los metadatos utilizan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento.

● **Confiabilidad:**

Los consumidores pueden confiar en todos los aspectos de los datos para la toma de decisiones. Se cuenta con un inventario de datos con capacidad para identificar automáticamente atributos similares, resolver ambigüedades y detectar relaciones con otros activos de datos. Los factores para evaluar la confiabilidad son:

- El linaje de los metadatos está bien definido a lo largo de su ciclo de vida.
- Se ejecutan técnicas mediante gestión de metadatos para evaluar y mejorar la calidad de los datos.
- Los metadatos de los datos maestros y de referencia están priorizados.

● **Vinculación:**

Pone en evidencia las conexiones entre fuentes dispares y aprovecha las relaciones entre conjuntos de datos para reducir la posibilidad de duplicación de información. Los consumidores pueden aprovechar dichas relaciones para incrementar la capacidad de análisis de los datos. Los factores para evaluar la vinculación son:

- Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.
- Existen mapeos cruzados para diferentes vocabularios de metadatos.
- Los metadatos tienen asignado un identificador global único y persistente.

● **Interoperabilidad:**

Los metadatos son susceptibles de intercambio con herramientas de terceros; es decir, tienen la capacidad para que los datos o las herramientas de una aplicación no cooperantes puedan integrarse o

trabajar juntos con un esfuerzo mínimo (Wilkinson et al., 2016). Tanto los consumidores como los productores tienen una representación y comprensión de los datos. Los factores para evaluar la interoperabilidad son:

- El protocolo de intercambio es abierto, libre y de aplicación universal (estandarizado).
- Los metadatos utilizan vocabularios que siguen los principios de requerimientos.
- Los metadatos incluyen referencias cualificadas a otros metadatos.
- Los metadatos pueden integrarse usando medios tecnológicos que facilitan el envío automático de contenido.

● ***Seguridad:***

Los consumidores saben que los datos están protegidos contra el uso y la manipulación no autorizada o fraudulenta. Mediante la gestión de reglas y políticas se consolida la seguridad y la privacidad de cualquier conjunto de datos. Los factores para evaluar la seguridad son:

- Los metadatos se publican con una licencia de uso de datos clara y accesible.
- El protocolo permite un procedimiento de autenticación y autorización cuando sea necesario.
- Las restricciones de uso están definidas por niveles y según el rol.

Las respuestas a los requerimientos son variadas y tienen muchos matices. Los requerimientos tienen una dependencia dependiendo de que el consumidor pertenezca a un sistema interno o externo. Bajo esta óptica, en esta etapa no es recomendable generar un instrumento definitivo; si se cierra, no permite explotar los conceptos que se han planteado en esta dimensión. En otras fases posteriores de madurez del proyecto se buscará explotar formatos para la estandarización de requerimientos.

Finalmente, es importante determinar con carácter específico que el resultado de esta actividad es una herramienta que permite priorizar los requerimientos a nivel de tipología y de categoría de metadatos, fundamentado en los seis principios y siete metas cuya articulación permite conectar un conjunto de datos cualquiera con aplicaciones tanto dentro como fuera de la Universidad. Entender los requerimientos de los metadatos ayudará a aclarar y a identificar los posibles obstáculos para definir la estrategia de metadatos.

7.3.3.2. Definir la estrategia de metadatos

La estrategia busca generar una cultura a nivel institucional alrededor de los metadatos y, como cualquier estrategia, debe tener una planificación que debe ser socializada con los miembros de las dependencias que tienen bajo su responsabilidad sistemas de información. En últimas, ellos son una pieza clave porque tienen un amplio conocimiento sobre la estructura, el lenguaje y las reglas de negocio; sin sus aportes no sería posible conocer el estado actual ni concebir soluciones para generar y mantener buenas prácticas en el futuro. Los pasos para establecer la estrategia son los siguientes: política de gestión de metadatos, identificar fuentes de metadatos, entrevistas, establecer una base de conocimiento de negocio y técnico, búsqueda de estándares, documentación y, medir el uso y la eficacia de la estrategia.

7.3.3.2.1. Política de gestión de metadatos

El equipo de gestión de la información debe definir metas a corto y largo plazo que deben estar alineadas con la política de gestión de metadatos porque las controversias surgen cuando los metadatos están incompletos o no existen. En general, los metadatos deben defender las decisiones políticas basadas en los datos.

El gobierno y la gestión es un camino de doble vía dentro de la estrategia porque es difícil llevar a cabo una gobernanza con éxito sin un entorno de metadatos gestionado, y a su vez, es difícil crear un enfoque sólido de gestión de metadatos sin gobernanza y/o administración.

7.3.3.2.2. Identificar fuentes de metadatos

No es un secreto que los metadatos a menudo se crean como un subproducto del procesamiento de la aplicación en los sistemas y no como un producto final, es decir, nunca se crean pensando en su consumo y explotación. Por tanto, las fuentes de metadatos pueden ser variadas: sistemas transaccionales, almacenes de datos, fuentes no estructuradas tales como documentos, correo electrónico, logs, normativas, entre otras. Por lo anterior, para optimizar el trabajo de esta actividad con múltiples fuentes de metadatos debemos apuntar a aquellas que privilegian datos maestros, de referencia, estandarizados, que están integrados con herramientas tecnológicas para su gestión y/o que apoyan la existencia de los macroprocesos de la Universidad.

7.3.3.2.3. Entrevistas

Con el objetivo de establecer una base de conocimiento técnico y de negocio es necesario hacer entrevistas. Este es un trabajo que se ha venido recopilando mediante el uso de Formularios de recolección y validación (**Anexo 5**).

La necesidad de hacer entrevistas para entender los datos de los sistemas pone en evidencia que existen metadatos humanos; es decir, que gran parte de los metadatos empresariales y de la historia de los sistemas de la Universidad existen en las cabezas de los empleados. Como producto de las entrevistas, estos metadatos humanos quedarán consignados en un formato electrónico para ser compartidos en el futuro con consumidores interesados.

7.3.3.2.4. Establecer una base de conocimiento de negocio y técnica

Validar, socializar y priorizar los resultados de las entrevistas en términos de metadatos. La implementación de este paso debe ser gradual y se debe apoyar con una herramienta tecnológica, sin embargo, esto debe ser prioritario en fases posteriores del modelo de gestión de la información.

7.3.3.2.5. Búsqueda de estándares

Para la búsqueda de estándares de metadatos nacionales e internacionales se recomienda privilegiar la adaptación sobre la adopción, esto es, los metadatos no vienen en tallas únicas, la aplicación de un estándar estará sujeta a la realidad y contexto de la Universidad. Esta actividad debe ser concertada porque tiene impacto en varias herramientas tecnológicas y varios macroprocesos. Debe estar alineada con las políticas y metas institucionales.

Seleccionar y definir un estándar de metadatos representa un riesgo durante la implementación de la estrategia porque existen metadatos que pertenecen a elementos muy específicos de cada sistema que no tienen y no deberían tener asociado un estándar internacional o nacional. Por ejemplo, el gobierno sobre la codificación de los apoyos socioeconómicos de Bienestar Universitario es exclusivo de esta dependencia, así que cualquier usuario que quiera hacer uso de ese conjunto debería acogerse a la codificación de SIBU.

Uno de los errores más comunes es privilegiar la elección de un estándar internacional sobre uno nacional. Ilustremos esta situación con un ejemplo: supongamos que usamos como estándar la codificación definida por *GeoNames*⁷, cuya base de datos geográficos abierta, abarca todos los países y contiene más de once millones de nombres de lugares. Sin embargo, haciendo un estudio detallado del sistema se encuentra que los datos de latitud y longitud son diferentes a los reportados por la fuente nacional. Además, una de las fuentes que usa *GeoNames* para alimentar su portal es el mismo DANE (**Tabla 25**). Es decir que finalmente la fuente primaria es el DANE y por tanto su codificación DIVIPOLA es la oficial. Frente a lo anterior se han identificado múltiples riesgos:

- El sistema *GeoNames* funciona de manera similar a *Wikipedia*, lo que implica que cualquier usuario registrado puede actualizar para bien o para mal los datos.
- *GeoNames* depende de los cambios definidos por el DANE y, por tanto, de la codificación DIVIPOLA.
- La codificación de *GeoNames* no es integrable con aplicaciones que ya hagan uso de la codificación DIVIPOLA.
- *GeoNames* no cuenta con una codificación para los municipios y/o los centros poblados.

Tabla 25 Comparativo de fuentes

Estándar	Código departamento	Departamento	Latitud	Longitud	Fuentes
GeoNames	33	Cundinamarca	5	-74.16667	DANE , IDECA , GeoMedellín
DIVIPOLA	25	CUNDINAMARCA	4.77112	-74.43111	DANE

7.3.3.2.6. Documentación

Este paso busca determinar métodos para consolidar y sincronizar los metadatos de múltiples fuentes. Uno de los retos de la documentación de metadatos es construir una solución sobre una plataforma de

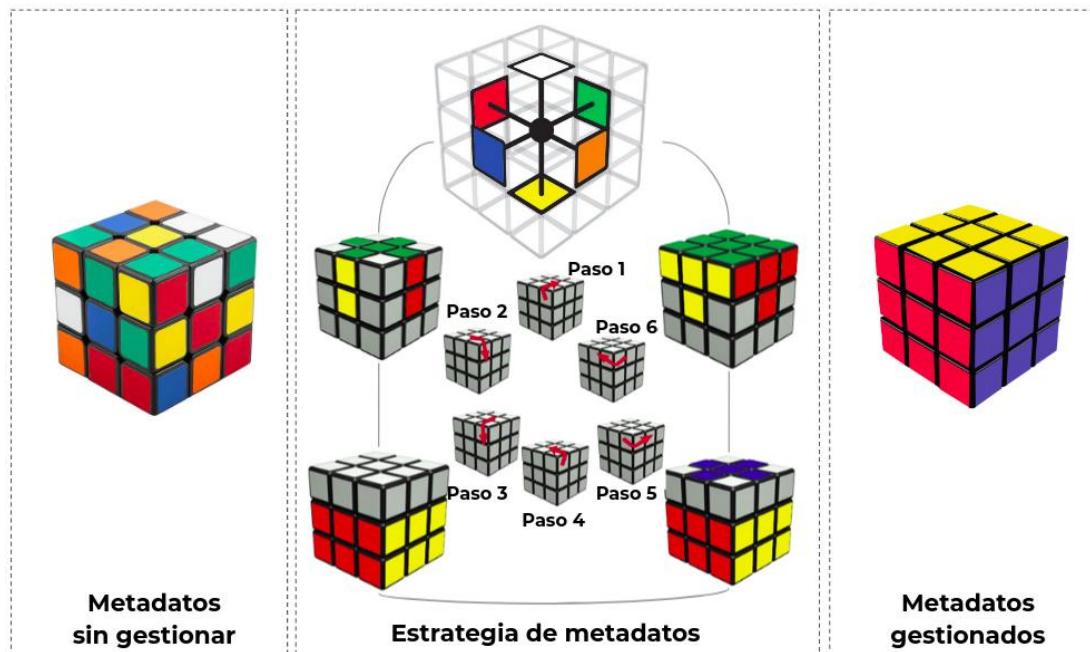
⁷ <https://www.geonames.org/about.html>

integración versátil que exponga sus propios metadatos para su consolidación y ofrezca la capacidad de integración para extraer metadatos de una amplia variedad de fuentes de datos dispares (Pervasive Software Inc., 2009). La primera consideración es extraer los metadatos de las fuentes aisladas y reunirlos en un punto central llamado registro de metadatos, MDR (por sus siglas en inglés). Esto ayudará a reducir la redundancia, la duplicación y las incoherencias de los metadatos. La segunda consideración, sincronización automatizada, garantiza que el registro central de metadatos se actualice con regularidad y permanezca sincronizado a lo largo del tiempo, aun cuando las fuentes sufran cambios o se añadan nuevas fuentes.

7.3.3.2.7. Medir el uso y la eficacia de la estrategia

Una estrategia de metadatos puede proporcionar una hoja de ruta para gestionar metadatos. Una estrategia de metadatos pone de manifiesto la importancia de un gobierno de metadatos para la Universidad que se preocupa por la calidad, la integridad y la reutilización de los datos. En el mediano plazo, el desarrollo y la implementación de una estrategia de metadatos permite a una organización comenzar a medir el valor de los datos bajo su control (Smith, 1997). Piénsese en un cubo de Rubik® (**Figura 33**), sin una estrategia, una persona podría emplear mucho tiempo intentando resolver el cubo y, al final, incluso después de varias iteraciones, no llegar a una solución. Ocurre lo mismo con la gestión de metadatos, una organización podría pasar años iterando sobre selección de estándares, adquisición de nuevas herramientas y aun así no alcanzar un nivel de madurez en su gestión. Lo mejor siempre será definir una serie de pasos que le permitan medir su avance, que este sea progresivo y que le ofrezca una retroalimentación para mejorar en cada iteración.

Figura 33 Analogía de estrategia de metadatos vs. Cubo Rubik



Si una estrategia no puede ser medida, entonces no sirve; en el caso del cubo, cada movimiento apunta a completar un segmento de los colores de ciertas caras. Si después de muchas vueltas no se consigue el patrón de colores deseado, algo debe estar fallando. Así mismo, la estrategia busca llevar el estado actual a una versión mejorada; en este caso, para definir cuál es el siguiente nivel, es conveniente definir niveles de madurez para la estrategia (Sulaiman, 2012), en particular los cinco siguientes (ver Figura 34).

1. **Inicial:** la Universidad tiene cierta sensibilización sobre la gestión de metadatos. Sin embargo, no dispone de herramientas o metodologías. Existen sistemas con deficiencias en la gestión de metadatos, lo cual genera malas interpretaciones sobre los datos y retrasos para crear nuevas aplicaciones.
2. **Gestionado:** los metadatos cumplen requerimientos, son estandarizados y están documentados por cada sistema de información. Se reconoce la procedencia de aquellos metadatos que están disponibles y que son útiles.
3. **Definido:** existe una política para metadatos y estos son transversalmente consistentes entre los múltiples sistemas de la Universidad.
4. **Medible:** los metadatos están centralizados y son utilizados para generar métricas de calidad.
5. **Optimizado:** los metadatos están soportados por herramientas tecnológicas completamente automatizadas para todos los sistemas de información.

Figura 34 Niveles de madurez para la estrategia de metadatos



Esta estrategia no es estática y evolucionará con el tiempo conforme se entiendan las implicaciones de los metadatos y el avance de las demás dimensiones en esta fase del proyecto. Asimismo, cambiará en la medida en que se fortalezca, se profundice y se amplíe el impacto de la DNED.

7.3.3.3. Proponer instrumentos de documentación de metadatos

Esta no es una tarea fácil debido a las diferencias en los formatos y a las áreas de conocimiento de los conjuntos de datos (Park, 2012). Los requerimientos y características de los metadatos en la Universidad hacen que irremediablemente su documentación se ciña a estándares seleccionados y que estos sean modificados apropiadamente para reflejar la realidad de los conjuntos de datos.

En su esencia, los instrumentos para la documentación de metadatos proporcionan una estructura en un formato fiable y predecible para la interpretación, una descripción resumida y uniforme, y sirven para describir los metadatos en cuatro aspectos (DataONE, 2016):

1. Términos comunes en lenguaje natural para permitir la coherencia entre los registros.
2. Definiciones comunes para facilitar la interpretación.
3. Un lenguaje común para facilitar la comunicación.
4. Estructura común para localizar rápidamente la información.

Hasta este punto es claro que la documentación de metadatos requiere de un rigor similar al de los datos. Después de todo, los metadatos también son datos, por tanto, pueden ser almacenados en una base de datos y pueden ser descritos, organizados y presentados a través del uso de un modelo (metamodelo). El metamodelo busca una representación que sea legible y entendible, es decir que está pensado para ser leído y entendido por personas. En esta primera fase se propuso adaptar el metamodelo definido en el estándar ISO/IEC 11179 - parte 3 porque define una estructura que es independiente del ambiente de aplicación, de la arquitectura, de los mensajes de intercambio de datos y del motor de base de datos. Cabe señalar que la norma fue especialmente concebida para el intercambio de datos basado en metadatos en entornos heterogéneos, asegurando definiciones exactas de los datos.

La norma ISO/IEC 11179 es una de las pocas normas consolidadas para el almacenamiento de metadatos empresariales en un entorno controlado. Cuando los elementos de datos están bien documentados según la norma y la documentación se gestiona en un registro de metadatos, se facilita la búsqueda y recuperación de los datos en bases de datos dispares, así como su envío y recepción a través de un proceso de intercambio (Reed, 2007). El núcleo del estándar es el registro de metadatos, MDR por sus siglas en inglés, y cumple tres objetivos principales: identificación, procedencia y control de calidad. La identificación se consigue asignando un identificador único a cada objeto registrado. La procedencia se refiere a la fuente de los metadatos y al objeto descrito. El control de calidad garantiza que los metadatos cumplan su función (ISO, 2015b). La finalidad de la norma ISO/IEC 11179 es promover:

- La descripción estándar de los datos.
- La comprensión común, armonización y normalización de los datos entre las personas dentro de la organización y entre terceros.
- La reutilización y normalización de los datos a lo largo del tiempo, el espacio y las aplicaciones.

- La gestión de los componentes de las descripciones de los datos.
- La reutilización de los componentes de las descripciones de datos.

Para familiarizarse con la norma, lo primero que se debe hacer es entender su vocabulario. En la **Tabla 26** se consignan las definiciones más importantes.

Tabla 26 Definiciones términos clave ISO 11179 parte 3

Término ⁸	Definición ISO	Ejemplo
Clase de objeto	Conjunto de ideas, abstracciones o cosas del mundo real que se identifican con límites y significado explícitos y cuyas propiedades y comportamiento siguen las mismas reglas	Persona
Propiedad	Una característica común a todos los miembros de una clase de objeto	Nombre, sexo, tipo de documento, fecha de nacimiento
Concepto de elemento de datos (DEC)	Una idea que puede representarse en forma de elemento de datos, descrita independientemente de cualquier representación particular (unión de una clase de objeto y una propiedad)	Nombre de la persona Sexo de la persona Tipo de documento de la persona Fecha de nacimiento de la persona
Dominio de valor (VD)	Conjunto de atributos que describen las características de representación de los datos de instancia con o sin valores permitidos enumerados	Sexo: {M = Masculino, F = Femenino, D = Desconocido} Tipo de documento: {CC = Cédula de ciudadanía, TI = tarjeta de identidad, CE = cédula de extranjería}
Elemento de datos (CDE)	Unidad de datos cuya definición, identificación, representación y valores admisibles se especifican mediante un conjunto de atributos (unión de un concepto de elemento de datos y un dominio de valor)	Tipo de documento de la persona - CC = cédula de ciudadanía

Para aplicar el estándar se empieza por verificar las preguntas y las respuestas de los formularios diligenciados por los diferentes sistemas de la información para determinar que se puedan agrupar en

⁸ <https://wiki.nci.nih.gov/display/caDSR/ISO+11179+Term+Definitions>

dos grandes conjuntos: conceptual y almacenamiento. El primero está asociado con aquellos metadatos que apuntan a responder sobre qué queremos recoger información (concepto de elemento de datos) y la segunda, todo lo que tiene que ver con su representación (dominio de valor).

La parte más básica y necesaria que hace eficaz la norma ISO es garantizar que se cumplan los requisitos mínimos de metadatos en todos los tipos de registros ya que permiten cierta flexibilización del registro de metadatos porque, según las necesidades del conjunto de datos, se pueden agregar nuevos atributos que adopten los principios y metas de los requerimientos.

En la **Tabla 27** se hace distinción de aquellos atributos de la norma que son obligatorios. Sin embargo, para futuras fases se podría hacer distinción entre aquellos atributos que son condicionales u opcionales. Hasta este punto el lector debe reconocer la importancia de atributos como definición y/o descripción. Además, es evidente que en el dominio de valor debe estar un metadato “tipo” (por ejemplo, VARCHAR, INT, STRING) que le proporciona a un desarrollador el suficiente contexto para no hacer inferencias que luego impliquen una falla de diseño en la implementación.

Tabla 27 Atributos mínimos para elementos de metadatos según la norma ISO 11179 parte 3

Concepto de elementos de datos	Objeto de clase
	Contexto (Sistema de información)
	Identificador único
	Propiedad
	Nombre extendido
	Abreviación
	Definición
	Versión
	Fecha de actualización
	Contacto
Dominio de valor	Nota de cambio
	Tipo de dominio {Enumerado, no enumerado}
	Tipo de dato
	Valores permitidos
	Descripción
	Unidad de medida
	Min/máx. longitud
	Rango (valores válidos)
	Fecha de actualización
	Contacto
	Nota de cambio

7.3.4. Salidas

De acuerdo con el alcance establecido, los artefactos producidos en esta fase son los que se proponen para recoger y documentar los metadatos. Adicionalmente se cuenta con un documento guía para **Código: U.MD.17.001.001** **Versión: 0.0** **Página 130 de 171**

diligenciar cada una de las secciones que componen el artefacto y, para finalizar, un documento que muestra el propósito de cada una de estas secciones. Estos artefactos se encuentran en el **Anexo 11**, así:

- Anexo 11: Formato para documentación de metadatos
- Anexo 11A: Instructivo de diligenciamiento de formato de documentación de metadatos
- Anexo 11B: Propósito de formato de documentación de metadatos

7.3.5. Usuarios

Los usuarios que pueden aprovechar la gestión de metadatos son aquellos que deben tomar decisiones sobre el propósito y el uso de los datos, así como en aquellos donde la estandarización de datos impacta en la calidad y la integración de estos, por ejemplo:

- Desarrollar nuevas aplicaciones, dado que deben definirse cuáles son los datos mínimos requeridos para la operación del sistema y en caso de realizar integraciones, establecer su estandarización.
- Administrar el repositorio de metadatos, dado que se utilizan los elementos mínimos requeridos que responden a las necesidades establecidas.
- Gestionar productos y servicios de información, dado que se deben establecer de forma adecuada su concepto, contexto, uso y la estandarización de algunos datos para presentar resultados que puedan integrarse a otras fuentes.

En general, los equipos técnicos y funcionales son usuarios de los metadatos, por cuanto representan una herramienta para dar una contextualización y usabilidad a los datos, permitiendo así mejorar en la interpretación de productos o servicios de la información que ofrecen.

8. CALIDAD DE DATOS

8.1. Conceptualización

8.1.1. Definición de la calidad de datos

La calidad de datos es el proceso de acondicionamiento de los datos para que satisfagan las necesidades de los usuarios de la comunidad Universitaria, tanto internos como externos. DAMA Internacional (2017) señala la calidad de datos como un programa de ejecución continua y no como un producto, ni como un proyecto. En la norma internacional ISO/IEC 25000 (Portal ISO 25000, 2005) se define la calidad como el grado en que las características de los datos satisfacen un conjunto de necesidades bajo unas condiciones específicas.

Según la recomendación de DAMA, los programas de calidad deben tener los siguientes principios:

- **Criticidad:** para identificar prioridades de mejora según la importancia de los datos
- **Gestión del ciclo de vida:** para facilitar la gestión en cada una de las etapas del ciclo de vida del dato, desde la creación o adquisición hasta la eliminación.
- **Prevención:** donde se identifiquen y mitiguen los riesgos de calidad de los datos.
- **Remediación de la causa raíz:** para mejorar la causa principal que da origen a riesgos en la calidad.
- **Gobierno:** corresponde a la definición y ejecución del programa de calidad de datos, que debe respaldar y mantener un entorno de datos gobernado.
- **Basado en estándares:** hace referencia a la definición y aplicación de estándares en los requerimientos de información.
- **Medición objetiva y transparente:** se refiere a que la metodología a aplicar debe ser de conocimiento de las partes interesadas en el proceso.
- **Integrado en procesos comerciales:** hace alusión al cumplimiento de estándares de la calidad de datos en todos los procesos.
- **Aplicación sistemática:** corresponde a la aplicación continua de los requerimientos de calidad de los datos.
- **Conectado a niveles de servicios:** se refiere a la conectividad de los reportes de calidad, la gestión de problemas, y los acuerdos de nivel de servicio.

De acuerdo con la metodología DAMA (2017), la actividad de gestión de la calidad de datos se enmarca en diez actividades que la preceden, generan insumos para su correcto desarrollo y conducen a desempeños exitosos en la gestión global de la información. Así, las empresas privadas o las entidades estatales que disponen de los insumos resultantes de las actividades se listan a continuación:

1. **Gobierno de Datos:** incluye las políticas de gobierno sobre los datos, cuyo fin es planificar, supervisar y controlar la gestión y el uso de datos.

2. **Gestión de Arquitectura de Datos:** corresponde al modelo de interoperabilidad sobre el que se gestionan los activos de datos.
3. **Desarrollo de datos:** se refiere a los estándares que se deben analizar, diseñar, implementar, probar, desplegar y mantener sobre los datos.
4. **Gestión de Operaciones de Datos:** proporciona el apoyo desde la adquisición del dato hasta su depuración.
5. **Gestión de la Seguridad de datos:** garantiza la privacidad, confidencialidad, acceso adecuado y demás políticas estatales que apliquen sobre el manejo de datos.
6. **Gestión de la Calidad de Datos:** consiste en los lineamientos que deben aplicarse para hacer seguimiento a la calidad del dato y mejorar su calidad.
7. **Gestión de Datos maestros y de referencia:** se refiere a los procesos para gestionar versiones originales y réplicas de los datos.
8. **Gestión de inteligencia de negocios y almacenamiento de datos:** consiste en el proceso de generación de informes y análisis con los datos.
9. **Gestión de contenidos y Documentación:** hace referencia a la gestión a ejecutar sobre los datos que se encuentran oficialmente fuera de las bases de datos internas.
10. **Gestión de metadatos:** es el proceso de integrar, controlar y proporcionar metadatos.

En la **Figura 35** se presenta la jerarquía establecida por DAMA (2017) para la administración y servicios de datos. La gestión de calidad de datos corresponde a un proceso en el que se monitorea el adecuado funcionamiento del sistema. Este monitoreo se realiza a través de la definición de indicadores que dan cuenta de la calidad de la información, dan respuesta a las reglas de negocio establecidas por cada dependencia de la entidad y ayudan en la medición de los niveles de servicio proporcionados por el sistema.

En el esquema de la figura se identifica como punto de partida el gobierno de datos (políticas y procedimientos para aplicar a los datos de la entidad) para el manejo de la información. Allí se planifica la gestión de datos, se hace la comprensión de las necesidades de la entidad, se establecen cuáles son las dependencias participantes, se identifican los roles de los administradores de datos, y otras características asociadas.

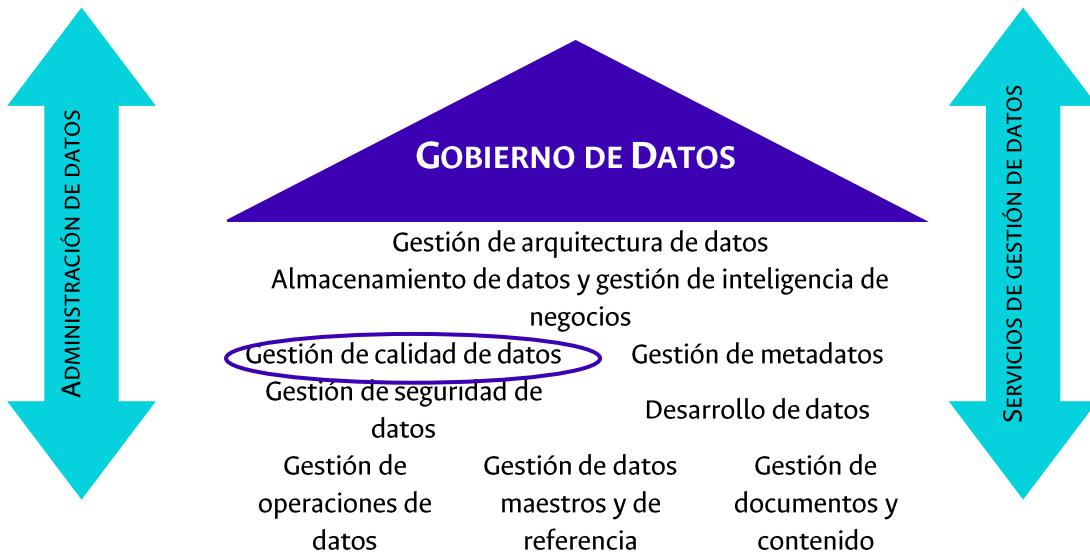
Como punto intermedio, dentro del esquema se ubica la arquitectura de datos, la cual debe contener modelos de datos conceptuales y lógicos de las áreas temáticas o dependencias de la Universidad, con sus reglas de negocio, propósitos, caracterización de las variables, relaciones entre las mismas, y demás características relacionadas⁹.

Posteriormente se identifican las características que deben estar presentes en la formulación y propuesta de los indicadores de calidad en particular para el momento de captura del dato o recepción

⁹ Vale la pena recordar que para este proyecto no se dispone aún de los lineamientos de gobierno y arquitectura de datos que acogerá la Universidad Nacional, ya que éstos se encuentran en construcción de manera paralela al modelo de gestión de la información

de la información de otra dependencia, entendiendo que este es el momento inicial del ciclo de vida del dato.

Figura 35 Gobierno de datos, administración y servicios. Guía de fundamentos para la gestión de datos (DAMA, 2010, p. 43)



El gobierno de datos en la Universidad Nacional de Colombia es un marco de trabajo integral que busca garantizar la eficiencia, confiabilidad y accesibilidad de la información para apoyar la toma de decisiones y el logro de los objetivos institucionales. En este contexto, la calidad de los datos juega un papel crucial, ya que asegura que los datos utilizados sean precisos, completos y consistentes a lo largo de toda la organización. La calidad de datos es esencial para garantizar la efectividad de las políticas y estrategias implementadas en la Universidad, así como para impulsar la innovación y el conocimiento a través de la investigación y la colaboración. Una alta calidad de datos permite a la Universidad mejorar significativamente la eficacia y eficiencia de sus procesos y operaciones.

Esto conduce a una mayor capacidad para tomar decisiones informadas, identificar oportunidades y abordar desafíos de manera efectiva. Además, al contar con información precisa y confiable, los diferentes sistemas de la información pueden operar de manera más integrada, facilitando la colaboración entre departamentos y la identificación de áreas de mejora. En última instancia un proceso adecuado genera un aumento en la reputación de la Universidad dado que facilita a la institución de decisiones asertivas para la sociedad. La medición de la calidad de datos es un componente esencial en el Gobierno de datos y existen diversas métricas que permiten evaluar distintos aspectos de la información. Alguna de estas permite tener una cuantificación clara de la situación actual de los datos.

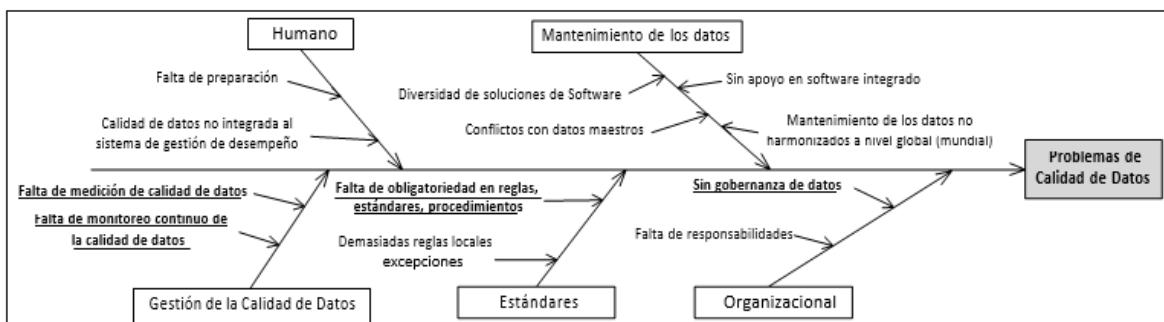
8.1.2. Características y metodología para la construcción de indicadores de calidad

El impacto de una inadecuada calidad en los datos se ve reflejado en una amplia gama de situaciones adversas para las organizaciones que se monetizan en pérdidas (desde financieras hasta reputacionales).

En el diagrama presentado en la **Figura 36** se identifican las principales causas que generan problemas en la calidad de la información; entre ellas, la falta de medición de la calidad y su monitoreo permanente.

Figura 36 Causas de problemas de calidad de datos – perspectiva organizacional corporativa.

Fuente: Rodríguez (2020, p. 7)



A partir de la exposición anterior se presentan los elementos para tener en cuenta en la construcción de indicadores que midan la calidad de los datos y su proceso de monitoreo. La Guía para Diseño, Construcción e Interpretación de Indicadores del DANE (2005) define un indicador como:

Una expresión cualitativa o cuantitativa observable, que permite describir características, comportamientos o fenómenos de la realidad a través de la evolución de una variable o el establecimiento de una relación entre variables, la que, comparada con períodos anteriores, productos similares o una meta o compromiso, permite evaluar el desempeño y su evolución en el tiempo. (p. 13)

La anterior definición de indicador fue llevada a la Guía de cuantificación, medición y seguimiento a la gestión de procesos (CMSPG) como “una expresión cualitativa o cuantitativa observable, que permite describir características, comportamientos o fenómenos de la realidad a través de la evolución de una variable o el establecimiento de una relación entre variables” (p. 11). El éxito en la formulación de indicadores está en elegir adecuadamente la información que mida y transmita el propósito misional de las dependencias de la Universidad Nacional de Colombia. Para cumplir con este propósito, la guía DANE señala que la medición debe tener en cuenta las siguientes características:

- **Ser pertinente**, es decir, que las mediciones que se lleven a cabo sean relevantes y útiles en el tiempo para facilitar las decisiones que serán tomadas sobre la base de tales mediciones.
- **Ser precisa**, es decir, que debe reflejar fielmente la magnitud que se quiere analizar, lo cual puede conseguirse a través de una buena definición de las características de las unidades a medir y una adecuada elección del instrumento de medición.
- **Ser oportuna**, es decir, que se encuentre disponible en el periodo de tiempo en que la información es importante y es relevante para la toma de decisiones, tanto para corregir como para prevenir.

- **Ser económica**, esto se refiere al hecho fundamental de que debe existir una proporcionalidad entre los costos incurridos en la medición y los beneficios y la relevancia de la medición (DANE, 2005, p. 12).

En la línea base de indicadores del DANE (2009) se señalan tres características básicas que debe cumplir todo indicador. En las siguientes secciones se detalla la clasificación y características que deben estar presentes en la formulación de indicadores:

1. **Simplificación**: la realidad en que se actúa es multidimensional y un indicador puede considerar una de tales dimensiones (económica, social, cultural, política, etc.) pero no puede abarcárlas todas.
2. **Medición**: permite comparar la situación actual de una dimensión de estudio en el tiempo o respecto a patrones establecidos.
3. **Comunicación**: todo indicador debe transmitir información para la toma de decisiones.

Además, los objetivos principales de los indicadores deben:

- Generar información útil para mejorar el proceso de toma de decisiones, el proceso de diseño, implementación o evaluación de un plan, programa, etc.
- Monitorear el cumplimiento de acuerdos y compromisos.
- Cuantificar los cambios en una situación que se considera problemática.
- Hacer seguimiento a los diferentes planes, programas y proyectos, para realizar los correctivos oportunos y mejorar la eficiencia y eficacia del proceso en general.

La Guía Internacional para la Gestión de Datos (DAMA Internacional, 2017) señala que “no existe un conjunto único y acordado de calidad [para los] datos” (p. 435); sin embargo, presenta la información de la **Tabla 28**, donde se describen las dimensiones de la calidad. De manera análoga, en la **Tabla 29** se encuentran definidas las características de calidad de la norma ISO/IEC 25012 (2019), donde se puede observar que ambas metodologías apuntan a características similares.

Tabla 28 Dimensiones comunes de la calidad de datos. Adaptado de: DAMA (2017, pp. 435-437)

Dimensión de calidad	Descripción
Exactitud	Grado en que los datos presentan correctamente entidades “de la vida real”. La exactitud es difícil de medir, a menos que una organización pueda reproducir colecciones de datos o confirmar manualmente la exactitud de los registros. La mayoría de las medidas de exactitud dependen de la comparación con una fuente de datos que ha sido verificada como exacta, como un sistema de registros o datos de una fuente confiable.
Compleitud	La completitud se pregunta si todos los datos están presentes. Esta puede medirse al nivel de un conjunto de datos, registro o columna (¿Contienen los datos todos los registros esperados? ¿Están los registros poblados correctamente? ¿Están las columnas/atributos poblados al nivel esperado?). Es posible asignar reglas de completitud a un conjunto de datos con distintos niveles de restricción: atributos

Dimensión de calidad	Descripción
	mandatorios que requieren un valor, elementos de datos con valores condicionales u opcionales, y valores de atributos inaplicables. Las mediciones a nivel de conjunto de datos pueden requerir la comparación a una fuente de registros o pueden estar basadas en niveles históricos de población.
Consistencia	La consistencia se refiere a asegurar que los valores de los datos sean representados consistentemente dentro de un conjunto de datos y entre conjuntos de datos, y consistentemente asociados entre conjuntos de datos. También se refiere al tamaño y composición de los conjuntos de datos entre sistemas a lo largo del tiempo. La consistencia puede estar definida entre los valores de un conjunto de datos y otro conjunto de atributos en el mismo registro (consistencia a nivel de registro), entre un conjunto de valores de atributos y otro conjunto de atributos en registros diferentes (consistencia entre registros), o entre un conjunto de valores de atributos y el mismo conjunto de atributos dentro del mismo registro en distintos momentos (consistencia temporal). La consistencia también puede ser utilizada para referirse a la consistencia de formato. No confundir la consistencia con precisión o correctitud.
Estandarización	Características que se espera sean consistentes dentro y entre conjuntos de datos pueden ser utilizadas como base para la estandarización de datos. La estandarización de datos se refiere al condicionamiento de la entrada de datos para asegurar que los datos cumplan las reglas de contenido y formato. Estandarizar datos permite realizar comparaciones efectivas y facilita la generación de salidas consistentes encapsula las restricciones de consistencia como un conjunto de reglas que especifican las relaciones consistentes entre valores de atributos, ya sea entre registros o mensajes, a lo largo de todos los valores de un solo atributo (como un rango o lista de valores válidos).
Integridad	Integridad de datos (o coherencia) incluye ideas asociadas con la completitud, exactitud y consistencia. La integridad se refiere usualmente a la integridad referencial (consistencia entre objetos de datos vía una clave referencial contenida en ambos objetos) o consistencia interna dentro de un conjunto de datos (que no existan vacíos o partes faltantes). Los conjuntos de datos sin integridad son vistos como corruptos o con pérdida de datos. Los conjuntos de datos sin integridad referencial tienen "huérfanos" -claves de referencia inválida o duplicados- registros idénticos que pueden afectar negativamente las funciones de agregación. El nivel de registro huérfanos puede ser medido como un conteo de registros o un porcentaje del conjunto de datos.
Razonabilidad	La razonabilidad pregunta si un patrón de datos cumple las expectativas. La medición de razonabilidad puede tomar diferentes formas. Por ejemplo, la razonabilidad puede estar basada en la comparación para generar data referencial, o instancias pasadas de conjuntos de datos similares. Algunas ideas acerca de la razonabilidad pueden ser percibidas como subjetivas. Si ese es el caso, es necesario trabajar con los consumidores de datos para articular la base de sus expectativas de datos para formular comparaciones objetivas. Una vez que las medidas

Dimensión de calidad	Descripción
	referenciales de razonabilidad están establecidas, pueden utilizarse para comparar objetivamente nuevas instancias de los mismos datos en el mismo conjunto de datos con la finalidad de identificar cambios.
Oportunidad	Este concepto se refiere a varias características de los datos. Las mediciones de oportunidad necesitan ser entendidas en términos de la volatilidad esperada (con qué frecuencia los datos cambian y por qué razones). La vigencia de los datos es la medida en la cual los valores de los datos son la versión más actualizada de la información. Datos relativamente estáticos pueden mantenerse vigentes por largos períodos de tiempo mientras que datos volátiles se mantienen vigentes por cortos períodos de tiempo. Algunos datos, como los valores de acciones en páginas web financieras, muestran generalmente mensajes que indican que los valores se tomaron en un momento del tiempo, de tal manera que los consumidores de datos entiendan que existe el riesgo de que los datos hayan cambiado desde que fueron registrados. La latencia mide el tiempo entre la creación del dato y cuando fue disponibilizado para su uso.
Unicidad	Establece que ninguna entidad existe más de una vez en el conjunto de datos. Confirmar la unicidad entre las entidades dentro de un conjunto de datos implica que el valor de la clave se relaciona con cada entidad única, y solamente esta entidad específica, dentro del conjunto de datos.
Validez	Se refiere a si los valores de los datos son consistentes dentro de un dominio de valores definido. Un dominio de valores puede ser un conjunto de valores válidos (como en una tabla de referencia), un rango de valores o un valor que puede ser determinado mediante reglas. El tipo de dato, formato y precisión de los valores esperados puede ser considerado en la definición del dominio. Los datos pueden también ser válidos solamente por un período específico de tiempo. Para validar los datos se necesita compararlos con las restricciones de dominio. Los datos pueden ser válidos (por ejemplo, pueden cumplir los requerimientos de dominio) y de todas formas no estar asociados con precisión o correctamente con registros particulares.

Tabla 29 Dimensiones de calidad según la norma ISO/IEC 25012. Fuente: Norma ISO/IEC 25012 (2019)

Dimensión (Característica)	Definición
Exactitud	<p>El grado en el cual el dato tiene atributos que representan el valor correcto de un concepto o evento en un contexto específico de uso.</p> <p>Se pueden distinguir dos tipos de exactitud: la sintáctica y la semántica.</p> <ul style="list-style-type: none"> La exactitud sintáctica se define como la cercanía de los valores de datos a un conjunto de valores definidos en un dominio considerado sintácticamente correcto.

Dimensión (Característica)	Definición
	<ul style="list-style-type: none"> La exactitud semántica se define como la cercanía de los valores de datos a un conjunto de valores definidos en un dominio considerado semánticamente correcto.
Compleitud	El grado en el cual el dato asociado a una entidad tiene valores para todos los atributos esperados e instancias de entidad relacionadas, de acuerdo con un contexto específico de uso.
Consistencia	El grado en el cual el dato tiene atributos libres de contradicción y son coherentes con otros datos en un contexto específico de uso.
Credibilidad	El grado en el cual el dato tiene atributos considerados como verdaderos y creíbles por usuarios en un contexto específico de uso.
Actualidad	El grado en el cual el dato tiene los atributos que son del período correcto en un contexto específico de uso.
Accesibilidad	El grado en el cual se puede acceder al dato en un contexto específico de uso.
Conformidad	El grado en el cual el dato tiene atributos que se adhieren a normas, convenciones o regulaciones vigentes y reglas relacionadas con la calidad de datos en un contexto específico de uso.
Confidencialidad	El grado en el cual el dato tiene los atributos que aseguran que éste es sólo accesible e interpretable por usuarios autorizados en un contexto específico de uso.
Eficiencia	El grado en el cual el dato tiene los atributos que pueden ser procesados y proporciona los niveles esperados de funcionamiento (desempeño) usando las cantidades y los tipos de recursos apropiados en un contexto específico de uso.
Precisión	El grado en el cual el dato tiene atributos que son exactos o que proporcionan su discriminación en un contexto específico de uso.
Trazabilidad	El grado en el cual el dato tiene atributos que proporcionan un rastro de auditoría de acceso a los datos y de cualquier cambio hecho a los datos en un contexto específico de uso.
Comprendibilidad	El grado en el cual el dato tiene atributos que le permiten ser leído e interpretado por usuarios, y es expresado en lenguajes apropiados, símbolos y unidades en un contexto específico de uso.
Disponibilidad	El grado en el cual el dato tiene atributos que le permiten ser recuperados por usuarios autorizados y/o aplicaciones en un contexto específico de uso.
Portabilidad	El grado en el cual los datos tienen atributos que les permiten ser instalados, substituidos o movidos de un sistema a otro conservando la calidad existente en un contexto específico de uso.
Recuperabilidad	El grado en el cual el dato tiene atributos que le permiten mantener y conservar un nivel especificado de operaciones y calidad, aún en caso de falla, en un contexto específico de uso.

8.1.3. Tipología de indicadores según la Guía de Cuantificación, Medición y Seguimiento a la Gestión de Procesos (CMSPG) de la Universidad Nacional de Colombia

En la Guía para Diseño, Construcción, e Interpretación de Indicadores DANE (2005), acogida en la Guía de CMSPG de la Universidad Nacional de Colombia se presenta una clasificación de indicadores que usamos a lo largo de este proyecto y tuvimos en cuenta para la construcción del producto final de la dimensión de Calidad¹⁰. La clasificación de indicadores es la siguiente:

- **Indicadores según medición:**

Según como se exprese su valoración pueden ser cuantitativos y cualitativos. Los indicadores cuantitativos son una representación numérica de la realidad. Su característica más importante es que, al encontrarse valores diferentes, estos pueden ordenarse de forma ascendente o descendente. Los indicadores cualitativos son otro instrumento que permite tener en cuenta la heterogeneidad, amenazas y oportunidades del entorno organizacional. Además, permiten evaluar, con un enfoque de planeación estratégica, la capacidad de gestión de la dirección y demás niveles de la organización. Su característica principal es que su resultado se refiere a una escala de cualidades.

- **Indicadores según nivel de intervención:**

Hacen referencia a la cadena lógica de intervención, es decir, a la relación entre los insumos, los resultados y los impactos. Los indicadores que se encuentran bajo esta clasificación tratan de medir en cuánto nos acercamos a las metas esperadas con los insumos disponibles. Para esto se dispone de cinco tipos de indicadores:

- *Indicadores de impacto:* se refieren a los efectos a mediano y largo plazo que pueden tener uno o más procesos, planes, programas, proyectos en el universo de atención y que repercuten en la sociedad en su conjunto.
- *Indicadores de resultado:* se refieren a los efectos de la acción institucional y/o de un plan, programa, proyecto sobre la sociedad.
- *Indicadores de producto:* se refieren a la cantidad y calidad de los bienes y servicios que se generan mediante las actividades de una institución o de un plan, programa, proyecto.
- *Indicadores de proceso:* se refieren al seguimiento de la realización de las actividades programadas, respecto a los recursos materiales, personal y/o presupuesto. Este tipo de indicadores describe el esfuerzo administrativo aplicado a los insumos para obtener los bienes y servicios programados.
- *Indicadores de insumo:* se refieren al seguimiento de todos los recursos disponibles y utilizados en una intervención.

¹⁰ Este producto es la formulación de indicadores, en la etapa inicial del ciclo de vida del dato, o captura de la información, junto con sus especificaciones de monitoreo.

- **Indicadores según jerarquía:**

- *Indicadores estratégicos:* estos indicadores permiten hacer una evaluación de productos, efectos e impactos; es decir, la forma, método, técnica, propuesta, solución y alternativa son elementos que pertenecen, bajo el criterio de estrategia, a todo el sistema de seguimiento y evaluación. En este sentido, los indicadores estratégicos permiten medir los temas de mayor incidencia e impacto.
- *Indicadores de gestión (o indicadores internos):* su función principal es medir el primer eslabón de la cadena lógica de intervención; es decir, la relación entre los insumos y los procesos.

- **Indicadores según calidad:**

Dan cuenta de la dinámica de actividades específicas. Estos indicadores deberán medir la eficiencia, eficacia y efectividad del proceso.

- *Indicadores de eficacia:* a través de ellos se controlan los resultados planeados, como: objetivos, metas, resultados de un plan, programa, proyecto o política, así como las características que satisfacen al usuario de los productos o servicios; por ejemplo: calidad, satisfacción, servicio, oportunidad de entrega, cantidad.
- *Indicadores de eficiencia:* hacen control de los recursos e insumos del proceso. Por ejemplo: consumos de materiales, gastos, productividad, desperdicios, reproceso, horas hombre asociadas al desperdicio.
- *Indicadores de efectividad:* involucran la eficiencia y la eficacia, es decir, el logro de los resultados programados en el tiempo con los costos más razonables posibles. Es la medida del impacto de nuestros productos en la sociedad o en las partes interesadas.

8.1.4. Criterios o características de los indicadores según la Guía de CMSGP de la Universidad Nacional de Colombia

A partir de la Guía para diseño, construcción, e Interpretación de Indicadores DANE (2005) acogida en la Guía de CMSGP de la Universidad Nacional de Colombia, se presentan los criterios o características que deben estar presentes al momento de proponer y construir un indicador.

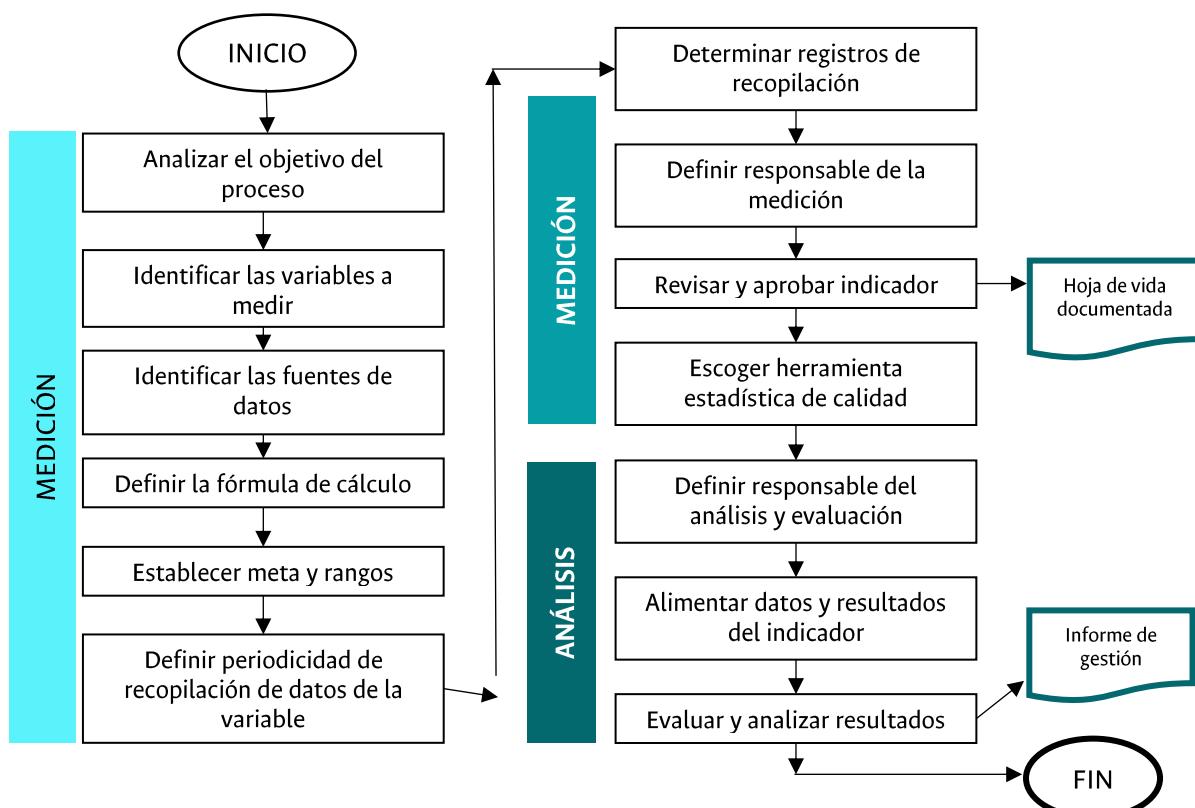
- **Especifico:** los indicadores deben estar orientados al objetivo que se quiere medir, deben ser objetivos y no estar condicionados por factores externos.
- **Medible:** los indicadores deben ser cuantificados, observados, analizados, probados y comparados.
- **Realizable:** los indicadores deben ser calculables (que sus cálculos se puedan realizar) y viables, en el sentido de que los costos de su cálculo sean inferiores a su utilidad. Además, sus patrones de comparación deben ser realistas, es decir, si se establecen metas o compromisos a ser medidos por el indicador, estas deben ser posibles de alcanzar o cumplir.

- **Relevante:** los indicadores deben presentar información valida sobre lo importante y para lo importante, el control y la toma de decisiones, etc.
- **Tiempo de medición oportuno:** los indicadores deben ser recogidos y reportados a tiempo, es decir, deben ser oportunos y hacer referencia a periodos específicos.

8.1.5. Pasos metodológicos para la construcción de indicadores

La **Figura 37** presenta y adapta los pasos metodológicos para la construcción de indicadores que se encuentran en la Guía de CMSGP de la Universidad Nacional de Colombia y una descripción de cada uno de dichos pasos. Los pasos que corresponden a la elección de la herramienta estadística a utilizar y los análisis a realizar dependen de los recursos que se dispongan en cada caso, y no corresponde a esta dimensión hacer las respectivas aclaraciones.

Figura 37 Pasos metodológicos para la construcción de indicadores. Fuente: Universidad Nacional de Colombia (2012, p.19)



1. Analizar el objetivo del proceso:

El objetivo especifica el resultado esperado del proceso (producto o salida) a partir del cual se obtienen los factores críticos de éxito (aquellos elementos del proceso que es necesario mantener bajo control

para garantizar el logro del objetivo) así como las características o parámetros de eficiencia y/o eficacia (variables de resultado esperado del proceso).

2. Identificar las variables a medir:

Se debe iniciar por identificar qué se quiere medir. Los indicadores deben proporcionar información concreta acerca de dicho objeto. Una vez identificado lo que se quiere medir se precisa qué aspecto específico se desea evaluar. Es posible identificar las características o cualidades que conformarán el indicador; estas unidades de análisis son conocidas como variables y pueden modificarse o variar en el tiempo y en el espacio.

Del conjunto de variables asociadas al objeto de medición se deben escoger aquellas relacionadas con el aspecto específico a evaluar. Una vez identificadas, las variables deben ser definidas con la mayor rigurosidad posible asignándoles un sentido claro para evitar que se originen ambigüedades y discusiones sobre sus resultados. Así mismo, se debe tener claridad sobre quién y cómo produce dicha información para establecer criterios de confiabilidad. Los factores críticos sirven para definir la variable a medir que compondrá el indicador. Esto con el fin de establecer qué realmente mide la gestión del proceso. La variable puede ser cuantitativa o cualitativa.

Es necesario definir el nombre de las variables ya que éste hace referencia específica a las características más importantes de cada una. Debe ser único y guardar relación con la estructura de nombres de las demás variables involucradas en el análisis.

Es necesario definir el nombre de las variables ya que éste hace referencia específica a las características más importantes de cada una. Debe ser único y guardar relación con la estructura de nombres de las demás variables involucradas en el análisis.

3. Identificar la fuente de los datos:

La fuente de los datos es el lugar en donde se puede verificar la existencia de los datos cualitativos o cuantitativos. Por ejemplo, sistemas de información, bases de datos, informes, archivos, encuestas y evaluaciones, registros de mantenimiento, registros, listas de verificación, listas de chequeo, observaciones, resultados de pruebas de inspección, promedios, porcentajes, desviaciones estándar, consolidados, rendimientos, consumos, pagos y otros. Cabe anotar que la fuente de los datos no es la entidad, dependencia, organización, etc., que produce la información, ni quien se encarga de reportarla, procesarla o consolidarla.

4. Definir la fórmula de cálculo:

La fórmula de cálculo es la expresión matemática que correlaciona el comportamiento de la variable con respecto a un logro esperado o planeado, esta nos dice cuánto se ha conseguido con respecto a lo que se planeó. La comparación de la variable puede ser a) un valor actual con respecto a un valor histórico (referencial), b) un valor real con respecto a un valor programado (referencial) y c) una parte con relación a un todo (referencial). Por lo tanto, con la fórmula matemática se calcula el indicador. La fórmula debe dar cuenta del nombre de la variable. Las siguientes son características que se deben tener en cuenta en el momento de definición de la fórmula de cálculo:

- *Nombre del indicador:* es la expresión verbal, precisa, única y concreta que identifica el indicador, debe ser corto y claro. Nota: Un indicador no realiza ninguna acción, es decir, no lleva verbos.
- *Tipo de indicador:* un indicador debe ser de fácil comprensión e interpretación. La **Tabla 30** presenta los criterios generales que existen para la selección de indicadores y tiene en cuenta las características de los datos que se van a utilizar como soporte.
- *Unidad de medida:* es el parámetro bajo el cual se expresa la variable. Es importante que quede establecido ya que pueden existir diferentes unidades de medida para una misma variable; por ejemplo, la edad puede ser expresada en años, en meses, en días, en años y meses, etc., dependiendo del contexto.
- *Tendencia:* es el tipo de trayectoria que se espera que tenga el indicador. Se define como la dirección ascendente o descendente y constante en el que se desempeña el indicador. La tendencia muestra el comportamiento del indicador a lo largo de un ciclo de tiempo. Permite comparar los avances reales frente al comportamiento programado.
- *Nivel de Aplicabilidad:* especifica si el indicador aplica a nivel de institucional, sede, facultad, centro, instituto o laboratorio.

Tabla 30 Criterios a valorar en la selección de indicadores. Universidad Nacional de Colombia (2012, p.22)

Criterio de selección	Pregunta para tener en cuenta	Objetivo
Pertinencia	¿El indicador expresa qué se quiere medir de forma clara y precisa?	Busca que el indicador permita describir la variable objeto de la medición
Funcionalidad	¿El indicador es monitoreable?	Verifica que el indicador sea medible, operable y sensible a los cambios registrados en la situación inicial
Disponibilidad	¿Los datos de la variable del indicador están disponibles?	Los indicadores deben ser construidos a partir de variables sobre las cuales se puedan obtener datos de tal manera que sean consultadas cuando sea necesario.
Confiabilidad	¿De dónde provienen los datos?	Los datos deben ser medidas siempre bajo estándares veraces, confiables y oportunos
Utilidad	¿El indicador es relevante con lo que se quiere medir?	Que los resultados y análisis permitan tomar decisiones

5. Establecer meta y rangos:

Establecimiento de la meta: la meta es el resultado aceptable que se espera del comportamiento de un indicador en un periodo de tiempo determinado y sirve para calificar su logro. Las metas deben ser reales

y alcanzables y dependen de la capacidad de la institución para lograrlas. Se pueden determinar de la siguiente forma:

- *A partir de datos históricos:* cuando se conocen los resultados pasados de un indicador, se puede realizar una proyección y determinar lo que se espera en el siguiente periodo.
- *A partir de comparaciones:* en muchos casos no se conocen las tendencias de indicadores que se han definido, pero son indicadores que en otras instituciones ya se han determinado y han logrado resultados. En este sentido, al observar una institución de igual complejidad se podrían tomar los datos del mismo indicador que se está estudiando y determinar una meta.
- *A partir de tendencias del entorno:* hay indicadores de los que se puede determinar una meta a partir de la información del entorno. Por ejemplo, el Ministerio de Educación constantemente arroja una serie de indicadores como el de estudiantes por docente del Sistema de Educación Superior y a partir de esta información se puede determinar la meta.
- *A partir de políticas institucionales o estatales:* las metas pueden ser definidas en función del cumplimiento de políticas. Por ejemplo, si se sabe que la meta según el Plan Nacional de Educación es aumentar en X porcentaje la tasa de graduación, esta información también sirve de base para fijar metas institucionales.
- *A partir de construcción participativa:* consiste en reunirse en equipo de trabajo y por medio de lluvia de ideas u otra técnica llegar a un acuerdo de una meta para un indicador determinado. Es más factible utilizar esta forma de definición de meta cuando no se tiene información de ningún tipo que sirva.
- A partir de la experiencia de quienes hacen parte del proceso.
- A partir de las estimaciones iniciales, en caso de no tener datos para efectuar el promedio.
- *A partir de la valoración de los recursos de que se dispone:* un indicador puede adoptar diversos valores a lo largo del tiempo, el valor inicial del indicador que se toma como referencia para comparar el avance del objetivo se llama línea base. La información que precisa la línea base es: -valor del indicador-año a que se refiere ese valor-período (se utiliza cuando en un año se realizan diversas mediciones). Si las metas se establecen para una administración semestral, la línea base recomendada es el valor del indicador al final de la administración anterior.

Establecimiento de los rangos: los rangos son los niveles de desempeño del indicador cuya valoración determinará el estado de riesgo o problema de la variable y dará origen a diferentes tipos de acciones de mejoramiento (acciones preventivas o correctivas). También se les conoce como Parámetros de semaforización. Los rangos pueden ser:

- Normal: cumplimiento satisfactorio de la meta.
- Riesgo: valores con tendencia al incumplimiento de la meta. Generan una acción preventiva.
- Problema: valores de incumplimiento de la meta. Generan una acción correctiva.

6. Definir periodicidad de recopilación de los datos de la variable:

La periodicidad hace referencia a la frecuencia con la que se calcula el indicador (diaria, semanal, quincenal, mensual, semestral, anual, etc.). Es necesario determinar todas las diferentes frecuencias en

las que la variable se calculará. Como regla general, la frecuencia debe ser la menor posible para que permita emitir alarmas oportunas y tomar decisiones para la mejora continua.

7. Determinar registro de recopilación:

Cada indicador debe poseer un medio de registro de la toma de los datos de la variable según la periodicidad establecida; este registro es la evidencia del cumplimiento, veracidad y confiabilidad de la toma de los datos.

8. Definir responsable de la medición:

Es la persona delegada por el proceso, área o dependencia para la alimentación de los datos y resultados del indicador.

9. Revisar y aprobar indicador:

El líder del proceso debe revisar y aprobar el o los indicadores determinados dejando evidencia de la realización de dicha actividad (acta, comunicado, oficio, etc.) y divulgar los indicadores establecidos.

10. Documentar la Hoja de Vida del Indicador

La hoja de vida del indicador es la herramienta clave en el ejercicio para establecer esquemas de medición más efectivos y útiles. Esta no es otra cosa que el detalle de la información pertinente de cada indicador formulado, precisando fuentes de la información, fórmula, variables, etc.

8.2. Propósito

La calidad de datos hace referencia al cumplimiento de los requisitos para el acondicionamiento de los datos de tal manera que satisfagan las necesidades de unos usuarios de la comunidad educativa de la Universidad Nacional de Colombia, tanto internos como externos. La dimensión de calidad se enmarca en la metodología DAMA (2017), a partir de la cual se definirá un conjunto de indicadores y especificaciones para medir la calidad de los datos en el momento inicial de su ciclo de vida; es decir, cuando el dato es capturado o recibido por alguna dependencia de la Universidad. A partir de dichos indicadores y especificaciones se establecen los lineamientos para monitorear de forma mensual y permanente la calidad de los datos en el momento inicial del ciclo de vida.

8.2.1. Metas

- Proponer un conjunto de indicadores y especificaciones que midan la calidad de los datos en el momento inicial del ciclo de vida del dato
- Definir los lineamientos para estimar los indicadores de calidad de los datos mensualmente, en el ciclo inicial de vida del dato.

8.2.2. Alcance fase II

El alcance de esta fase se orienta a formular la propuesta de indicadores, estándares y especificaciones que miden la calidad de los datos en el momento inicial del ciclo de vida del dato, de forma transversal a todas las dependencias de la Universidad, lo cual debe ser insumo para la política de calidad de los datos. Además, establece los lineamientos con los que, posterior al cálculo de los indicadores, se mida, monitoree y reporte la calidad de los datos.

Los lineamientos servirán en fases posteriores de este proyecto para diagnosticar el estado de la calidad de la información manejada en diferentes dependencias de la Universidad Nacional de Colombia. Dicho diagnóstico se logrará con la información que se recoja en la prueba piloto de la fase I de este proyecto, lo cual, a su vez, servirá para definir estrategias de calidad en torno al procesamiento de la información.

De igual manera, dentro del propósito global de calidad, debe considerarse la estimación de indicadores que den cuenta del nivel de satisfacción de los servicios prestados por la Universidad Nacional de Colombia en cada una de sus dependencias. No obstante, esto se desarrollará en fases posteriores junto con tableros de control para monitorear permanentemente las características de calidad.

8.3. Desarrollo de la fase II de calidad de datos

La **Tabla 31** esquematiza los elementos que intervienen en el desarrollo de la calidad de los datos. En primera instancia se encuentran los proveedores de la información, que corresponden a los sistemas que capturan los datos. En una etapa posterior, a estos proveedores se les deberá aplicar una estandarización y unos indicadores que midan la calidad, los cuales deben estar escritos en las políticas de la Universidad y responder a las necesidades de la comunidad.

Hoy en día se presentan los lineamientos de estas políticas, por lo cual es necesario definir qué es la calidad de los datos, y, a partir de esta definición, determinar unos indicadores y, por supuesto, unas reglas para monitorearlos de forma mensual. Eso último contribuye al trabajo de los analistas de información, tomadores de decisiones y consumidores de datos en general.

Debido a que hoy en día no hay políticas específicas para medir y monitorear la calidad de la información en la Universidad Nacional de Colombia, la propuesta de indicadores y condiciones de monitoreo constituyen las salidas de esta dimensión.

Tabla 31 SIPOC para el desarrollo de la fase II de calidad de datos

Proveedores	Entradas	Actividades	Salidas	Usuarios
DNED	Modelo de GI	Realizar una propuesta de indicadores de calidad para el momento inicial del ciclo de vida del dato	Primera propuesta de fichas técnicas de indicadores de calidad para el momento inicial del ciclo de vida del dato	Líderes, analistas y administradores de la información y de sistemas de información

DNED	Primera propuesta de fichas técnicas de indicadores de calidad para el momento inicial del ciclo de vida del dato	Socialización y ajuste de indicadores de calidad para el momento inicial del ciclo de vida del dato	Fichas técnicas de indicadores de calidad para el momento inicial del ciclo de vida del dato ajustados	Comité de Gobierno de la Información
DNED	Modelo de GI, Política de GI, repositorios de datos de la Universidad	Priorizar datos para realizar evaluación de calidad	Documento con priorización de datos para la realización de evaluación de calidad socializado	Comité de Gobierno de la Información.

8.3.1. Proveedores

Los proveedores de esta dimensión son los sistemas que capturan la información a través de las plataformas de las que dispone la Universidad en sus diferentes dependencias para registro y control de estudiantes o personal administrativo.

8.3.2. Entradas

Las entradas de esta dimensión corresponden a:

- Las políticas de la Universidad Nacional de Colombia sobre estandarización de los datos, cálculo de indicadores de calidad y procedimientos de monitoreo de la calidad de la información.
- Las necesidades de calidad y reglas de manejo de la información de cada dependencia de la Universidad.
- Las fuentes oficiales de la información y repositorios de datos de la Universidad.

8.3.3. Actividades

8.3.3.1. Propuesta de indicadores de calidad para el momento inicial del ciclo de vida del dato

Debido a las limitaciones de integración de plataformas que capturan la información en las diferentes dependencias de la Universidad, esta dimensión presenta una propuesta de indicadores para medir completitud, consistencia y unicidad de criterios y registros erróneos como línea base de la Universidad para el momento del nacimiento del ciclo de vida del dato.

Ahora, la línea base corresponde a la primera medición de todos los indicadores considerados en un proyecto. Ésta da a conocer el valor de los indicadores al momento de ejecutarse unas acciones proyectadas, con lo que se establece el estado o 'punto de partida' del nacimiento del ciclo de vida del dato. En la línea base se consideran las siguientes seis acciones para definir el conjunto de indicadores asociado al momento del nacimiento del dato.

- **Construcción de un diccionario de información:**

Corresponde al establecimiento de los atributos que describen el contenido de una variable. Estos atributos se presentan en la **Tabla 32** y serán construidos en fases posteriores de este proyecto a partir de los insumos recopilados en la prueba piloto.

Tabla 32 Características del diccionario de datos de cada dependencia. Fuente: construcción propia

CARACTERÍSTICA	DESCRIPCIÓN
Nombre Sistema	Nombre del sistema que almacena la base
Nombre Variable	Nombre de la variable
Es llave	Marcar con SI, en caso de que este sea el criterio para cruzar con otras bases de la información y NO en otro caso
Descripción de la variable	Indicar brevemente la información que se almacena
Naturaleza	Indicar si la variable es cualitativa o cuantitativa
Valores admisibles	Indicar los valores permitidos
Restricciones	Indicar las restricciones permitidas para que sea aceptado el valor ingresado
Longitud	Indicar el número de caracteres permitidos para el almacenamiento
Fuente	Indicar si la variable se captura manualmente, de forma automática, o la fuente si corresponde a información primaria capturada de otra dependencia. Si es de forma automática, indicar el procedimiento

- **Reglas de estandarización de los atributos:**

A partir de la información recopilada en la prueba piloto se identificarán cuáles son las variables llaves entre dependencias, y en general, se definirán las reglas de negocio. Esto con el fin de estandarizar las características que describen las variables y que éstas se manejen de forma uniforme en todas las dependencias de la Universidad.

Una forma de estandarización es la utilización de códigos asociados a cada concepto de la información y, eventualmente, la construcción de un diagrama de entidad relación con la información aportada por las diferentes dependencias de la Universidad en la prueba piloto. El procedimiento de estandarización debe estar alineado con el cumplimiento de políticas de datos abiertos establecidos en el país y con lineamientos o reglas de negocio internas de la Universidad.

La consecuencia de estandarizar es mitigar el riesgo de duplicidad y reprocesamiento de la información, lo que simultáneamente mejora los tiempos de respuesta a solicitudes de información, implícitamente mejora de la interoperabilidad y permite detectar fallas asociadas a la capacidad de funcionamiento de los sistemas.

- **Propuesta de indicadores de calidad como línea base o diagnóstico de calidad de los datos:**

En calidad se utiliza el concepto de métrica para indicar la manera en que el proceso de control de calidad medirá un trabajo o un producto. Según DAMA (2017) la métrica permite cuantificar y definir la variación entre lo observado y lo esperado. En este capítulo, las métricas a utilizar corresponden a conteos y proporciones. La **Tabla 33** y la **Tabla 34** presentan métricas de la investigación de metadatos estadísticos para el aseguramiento de la calidad de los datos (Rodríguez, 2020).

Tabla 33 Métricas básicas de propósito general para datos cuantitativos

Granularidad	ID	Métrica	Función de Medición	Descripción	Tipo (O/S)	Método
Variable	Ntvar	Número total de variables	ntvar= contar (variables)	Número de variables (o columnas) que existen en el archivo de datos	O	Contar frecuencias
Registro	Ntreg	Número total de registros	ntreg= contar (registros)	Número de registros en un archivo de datos	O	Contar frecuencias
Archivo de datos	Ntdato	Número total de datos	ntdato=nv*nf	Número total de datos del archive (tabla)	O	Producto
Variable	Max	Máximo valor	max (variable)	El valor máximo de los valores de datos	O	Estadístico
Variable	Min	Mínimo valor	min (variable)	El valor mínimo de los valores de datos	O	Estadístico
Variable	Media	Media	media (variable)	La media de los valores de datos	O	Estadístico
Variable	Moda	Moda	moda (variable)	La moda de los valore de datos	O	Estadístico
Variable	Dvs	Desviación Estándar	dvs (variable)	La desviación estándar calculada	O	Estadístico
Variable	Mediana	Mediana	mediana (variable)	La mediana de los valores de datos	O	Estadístico

Tabla 34 Métricas para dimensiones inherentes asociadas a datos cualitativos

Dimensión	Factor	Problema Calidad de Datos	Granularidad	ID	Métrica	
EXACTITUD	Sintáctico	Dato con error ortográfico	archivo de datos	Acc-D1	Dato correcto – ortografía	
		Violación de dominio	variable	Acc-D2	Dato válido	
	Semántica	Atípico	archivo de datos	Acc-D3	Dato en la distribución esperada	
		Violación de dependencia funcional	archivo de datos	Acc-D4	Dependencia funcional	
COMPLETITUD		Dato faltante	variable, registro, archivo de datos	Com-D1	Dato completo – válido	
		Dato ficticio	variable, registro, archivo de datos	Com-D2	Dato completo – real	
CONSISTENCIA		Duplicado	registro	Cons-D1	Registro único	

A partir de los insumos que se obtendrán en la prueba piloto, es factible estimar los siguientes indicadores de gestión de calidad que cumplen con los supuestos expuestos en la sección anterior. Estos deben ser implementados en un manual de manejo de datos, con el fin de que cada dependencia los estime de forma mensual, y programe acciones de mejora cuando corresponda.

- *Porcentaje de Completitud de la variable =*

$$1 - \frac{\text{Número de registros con información faltantes}}{\text{Número total de registros}}$$

- *Porcentaje de registros erróneos =*

$$\frac{\text{Número de registros con información no permitida}}{\text{Número total de registros}}$$

- *Porcentaje de información que no cumple el criterio de ingreso =*

$$\frac{\text{Número de datos que no cumplen las condiciones de almacenamiento de la variable}}{\text{Número total de registros}}$$

- *Porcentaje de Unicidad de criterios =*

$$\frac{\text{Número de registros coincidentes en características entre dos bases que contienen la misma variable}}{\text{Número máximo de registros entre las dos bases}}$$

- *Porcentaje de registros consistentes =*

$$\frac{\text{Número de registros coincidentes entre dos bases que contienen la misma variable}}{\text{Número máximo de registros entre las dos bases}}$$

- *Atributos con riesgo de inconsistencia =*

$$\begin{cases} 1 \text{ corresponde a la variable que resulte inferior al 80\% en 3.1 y 3.2} \\ 0 \text{ en otro caso} \end{cases}$$

Los anteriores indicadores responden, por un lado, a los problemas de calidad identificados en Rodríguez (2020) y presentados en la columna 3 de la **Figura 38** y, por otro lado, a la relación identificada entre cada problema de calidad y las dimensiones que afecta (**Tabla 35**).

Figura 38 Clasificación de problemas de calidad de datos. Fuente: Rodríguez (2020, p. 6)

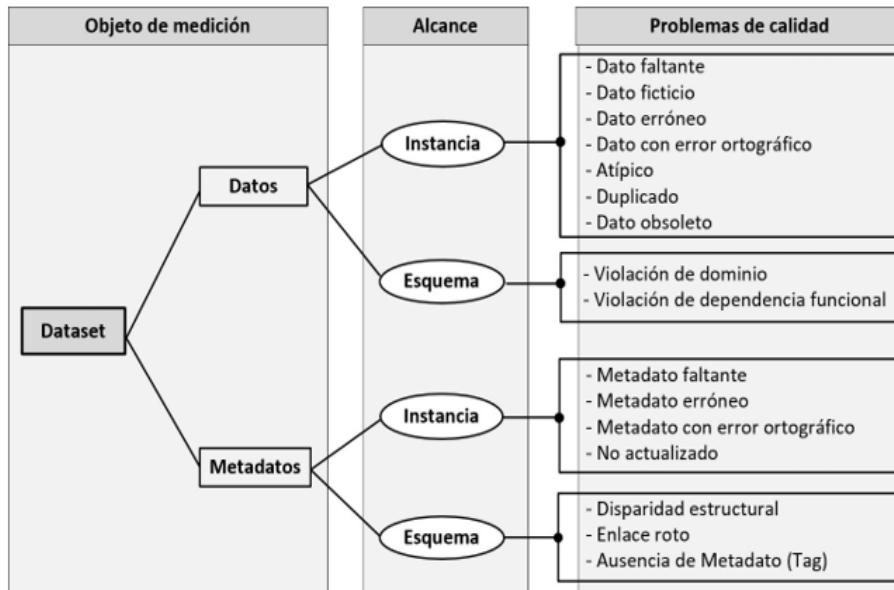


Tabla 35 Relación entre dimensiones y problemas de calidad – Objeto de medición: DATOS. Fuente: Rodríguez (2020, p. 65)

Problema de calidad	Dimensiones Inherentes				Dimensiones dependientes		
	Exactitud	Completitud	Consistencia	Rel. Tiempo	Accesibilidad	Confiabilidad	Relevancia
Instancia	Dato faltante		X			X	
	Dato ficticio		X			X	
	Dato Erróneo	X					
	Dato con error ortográfico	X					
	Atípico	X				X	
	Duplicado			X		X	
Esquema	Violación de dominio	X					
	Violación de dependencia funcional	X					

- Propuesta de reglas adicionales para el monitoreo de la calidad de la información:

Posterior a la fase I del proyecto de gestión de la información se definirá una matriz de riesgos para las diferentes dependencias de la Universidad, siempre que se disponga de los insumos requeridos para su cálculo, donde se registre información asociada a:

- Riesgo de pérdida de la información presente en cada sistema.
- Riesgo de rezago tecnológico presente en cada sistema.
- Nivel de seguridad de la información de cada sistema.
- Capacidad de los sistemas para generar respaldos de información.
- Registro de problemas de funcionamiento:

- *Fallas por caída del sistema (mensuales) =*

$$\frac{\text{Horas reportadas con fallas de funcionamiento en el sistema (mes)}}{\text{Total de horas del mes}}$$

- *Fallas en el sistema, según causas (mensual) =*

$$\left\{ \begin{array}{l} \frac{\text{No fallas asociadas a la causa 1}}{\text{Total de fallas reportadas en el mes}} \\ \frac{\text{No fallas asociadas a la causa 2}}{\text{Total de fallas reportadas en el mes}} \\ \dots \\ \frac{\text{No fallas asociadas a la causa } k}{\text{Total de fallas reportadas en el mes}} \end{array} \right.$$

- *Tiempo de fallas en el sistema, según causas (mensual) =*

$$\left\{ \begin{array}{l} \frac{\text{Suma de los tiempos de fallas asociadas a la causa 1}}{\text{Total de horas que falló el sistema en el mes}} \\ \frac{\text{Suma de los tiempos de fallas asociadas a la causa 2}}{\text{Total de horas que falló el sistema en el mes}} \\ \dots \\ \frac{\text{Suma de los tiempos de fallas asociadas a la causa } k}{\text{Total de horas que falló el sistema en el mes}} \end{array} \right.$$

- Capacitación de funcionarios respecto a obligaciones contractuales en el manejo de datos:

- *Porcentaje de asistencia a capacitaciones del sistema (trimestral) =*

$$\frac{\text{Número de asistentes invitados a capacitación(trimestre)}}{\text{Total de invitados a capacitación}}$$

- *Calificación cualitativa de la ejecución contractual trimestral:*

Para cada uno de los riesgos anteriormente considerados debe definirse una escala numérica; un listado de acciones a calificar; un responsable de la calificación; y unas acciones de mejora.

- **Perfilado de los datos:**

Teniendo como insumo la información que se recopile en las pruebas piloto de la fase II es posible definir un perfil de la actividad misional de cada dependencia donde se identifiquen unas problemáticas muy particulares que incluso den cuenta del nivel de satisfacción de los servicios prestados por la Universidad Nacional de Colombia, lo cual será desarrollado en fases posteriores.

- **Condiciones de seguimiento de los indicadores de calidad:**

A partir de los insumos anteriores, que se proyectan para una fase posterior a la II, es factible definir los lineamientos que deben ser contenidos en un manual de manejo de datos:

- Designación de responsabilidades y responsables, fuentes (en cuanto a la ejecución de tareas) y seguimiento a la calidad e interoperabilidad de la información. En particular, la designación de una oficina que administre el funcionamiento adecuado del flujo y calidad de la información, que podría ser la oficina de planeación y estadísticas de la Universidad.
- Definición de los tiempos para cortes de la información y generación de reportes.
- Incorporación del diccionario de la información anteriormente descrito.
- Incorporación de una bitácora que almacene la fecha de valoración de calidad, aplicación de procedimientos y resultados obtenidos.
- Introducción del registro de una bitácora de incidentes de calidad de la información.
- Inclusión de las políticas de integración entre sistemas de la información existentes y nuevos desarrollos, así como su socialización ante la comunidad universitaria.
- Identificación de las fuentes de la información oficiales.

8.3.3.2. Socialización y ajuste de indicadores de calidad para el momento inicial del ciclo de vida del dato

La Oficina de Gestión de la información de la DNED deberá encargarse de realizar la socialización y validación de los indicadores de calidad que han sido presentados en la sección anterior. La socialización deberá realizarse con los miembros permanentes previstos del Comité de Gobierno de la Información (Vicerrectoría General – SIGA, DNPE y DNED los cuales tienen competencias relacionadas con la adaptación de los datos a estándares de calidad dentro de la Política de Gestión de la Información. Las competencias de estos actores son las siguientes:

- Establecimiento de controles y verificaciones de calidad en compañía de gestores y custodios de datos.
- Definición y aplicación de normas técnicas nacionales e internacionales.
- Adaptación sistemática de estándares dentro de la operación y gestión de los sistemas de información.
- Producción de la información bajo estándares.

En este sentido, se deberá definir el procedimiento más conveniente para que los indicadores definidos en la sección anterior puedan ser conocidos por las dependencias que conformarán el Comité de

Gobierno de la Información, este procedimiento podrá consistir en reuniones bilaterales con la DNED o reuniones conjuntas entre las tres partes. Así mismo, se deberá valorar la conveniencia de invitar a estas reuniones a las dependencias de la universidad o equipos técnicos que sean los gestores y custodios de los datos que se deseen priorizar en el primer ejercicio de evaluación de calidad de los datos.

Junto con esta fase de socialización, la DNED deberá realizar los ajustes que se consideren pertinentes para estos indicadores teniendo en cuenta estándares nacionales e internacionales de calidad de los datos, así como las consideraciones que presenten las distintas partes. En este sentido, se deberá obtener como salida una versión ajustada de las fichas técnicas de los indicadores de calidad de los datos.

8.3.3.3. Priorizar datos para realizar evaluación de calidad

Teniendo en cuenta el extenso volumen de datos que es manejado en la Universidad, es necesario que los primeros ejercicios de evaluación de calidad de los datos se realicen partiendo de una priorización de los datos críticos de la Universidad (DAMA International, 2017, p. 451). El Comité de Gobierno de la Información o quien ejerza temporalmente el rol de gobierno de los datos en la Universidad deberá realizar esta priorización y elaborar un documento en el que esto quede consignado y que incluya las consideraciones técnicas, organizacionales y/o estratégicas por las cuales se ha optado por dicha priorización.

8.3.4. Salidas

Las salidas de este proceso corresponden a:

8.3.4.1. Formulación de indicadores que midan la calidad de los datos en la etapa inicial del su ciclo de vida, o captura de la información, junto con su instructivo de construcción

En la **Tabla 36** se presenta el listado de indicadores que se pueden aplicar de forma transversal a todas las dependencias de la Universidad Nacional para medir la calidad de la información en el momento inicial del ciclo de vida del dato.

Tabla 36 Listado de indicadores para aplicar transversalmente a las dependencias de la UNAL

	Nombre del indicador	Formula del indicador	Objetivo
1	Porcentaje de Completitud	$1 - \frac{\text{Número de registros con información faltantes}}{\text{Número total de registros}}$	Identificar el porcentaje de registros con información faltante.
2	Porcentaje de registros erróneos		Identificar el porcentaje de registros con información no permitida, y que resulta no consistente con la naturaleza

		<i>Número de registros con información no permitida</i> <i>Número total de registros</i>	de la variable medida.
3	Porcentaje de la información que no cumple el criterio de ingreso	$\frac{\text{Número de datos que no cumplen las condiciones de almacenamiento de la variable}}{\text{Número total de registros}}$	Identificar el porcentaje de registros que no cumplen con la regla de administración de ingreso de información.
4	Porcentaje de Unidad de criterios	$\frac{\text{Número de registros coincidentes en características, entre dos bases que contienen la misma variable}}{\text{Número máximo de registros entre las dos bases}}$	Identificar las variables comunes a bases de datos diferentes, que no manejan los mismos criterios de administración de la variable y por lo tanto rompen la definición de unicidad.
5	Porcentaje de registros consistentes	$\frac{\text{Número de registros coincidentes entre dos bases que contienen la misma variable}}{\text{Número máximo de registros entre las dos bases}}$	Identificar el porcentaje de registros coincidentes para las variables comunes a bases de datos diferentes.
6	Atributos con riesgo de inconsistencia por base de datos	Indicador=Suma de los 1/Total de registros a comparar, donde $\begin{cases} 1 \text{ corresponde a la variable que resulte inferior al } 80\% \text{ en 5} \\ 0 \text{ en otro caso} \end{cases}$	Identifica el número de variables con riesgo de inconsistencia, por base de datos.

Los indicadores anteriormente propuestos tienen las siguientes clasificaciones:

- **Según su nivel de medición:** son cuantitativos.
- **Según su nivel de intervención:** son de producto ya que miden la calidad de la información que administra cada sistema de la información de la Universidad, y a la vez son de proceso, porque monitorean a los estudiantes, administrativos, y en general a los recursos de la Universidad.
- **Según su nivel de jerarquía:** son de gestión, debido a que miden el primer nivel de intervención de la Universidad.
- **Según su nivel de calidad:** son de eficacia, ya que el principal objetivo es controlar la información que se recopila.

8.3.4.2. Ficha técnica de los indicadores de calidad de los datos en la etapa inicial de su ciclo de vida

Las características generales que debe contener la ficha técnica de cada indicador propuesto. Esto se hace de forma genérica debido a que: a) el recorrido permitido depende de las variables puntuales de cada sistema o dependencia de la Universidad, b) los criterios de aceptación pueden ser subjetivos y determinados por cada dependencia, mas no su sentido, c) la meta depende de las validaciones aplicadas en cada sistema o dependencia para capturar la información, y en la medida en que estas validaciones sean más estrictas, la meta será más exigente.

8.3.4.3. Procedimiento para la implementación de los indicadores de calidad propuestos

Teniendo en cuenta que las diferentes dependencias de la Universidad manejan información asociada a procesos misionales diversos, a continuación, se presenta un instructivo genérico para la implementación y adopción de indicadores de calidad.

1. Disponer de un listado de las variables contenidas en el sistema o dependencia, que incluya, el nombre, una descripción, la naturaleza (cualitativa o cuantitativa), y los valores permitidos.
2. De acuerdo con la descripción y naturaleza de la variable, se debe establecer el criterio de validación, es decir, precisar sobre las categorías o recorrido permitido en cada variable.
3. Identificar las variables sobre las cuales interesa monitorear la calidad de la información bajo los indicadores propuestos, ya que algunas variables corresponden a identificadores o almacenan datos que no son prioritarios o directamente relacionados con la misión de la dependencia, y en ese sentido no requiere la implementación y monitoreo de indicadores de calidad.
4. Indicar para cuáles variables se calculará cada indicador de calidad.
5. Generar un reporte donde se identifiquen los hallazgos de calidad en cuanto al no cumplimiento de los criterios de validación establecidos en el paso No. 3.
6. Llenar las fichas técnicas de cada indicador. Lo que implica definir cada uno de los atributos allí establecidos.
7. Incluir indicadores adicionales de calidad, que se consideren pertinentes, de acuerdo con la naturaleza misional de cada dependencia, y la disponibilidad de la información correspondiente.
8. Proponer salidas de la información visuales de los indicadores, de acuerdo con los niveles de desagregación establecidos en la ficha técnica.
9. Proponer la intervención que se requiera para mejorar el estado de calidad de la información.

9. CONCLUSIONES Y RECOMENDACIONES

En la fase I de este proyecto se trabajó sobre la definición de la calidad de datos y la propuesta de indicadores de completitud, consistencia, unicidad de criterios y registros erróneos como línea base de la Universidad para el momento de nacimiento del ciclo de vida del dato. Debido a las limitaciones de integración de plataformas que capturan la información en las diferentes dependencias de la Universidad se limitó la propuesta de indicadores.

Para el desarrollo de este documento, la definición de calidad de datos se extrajo de DAMA International (2017) que la define como un proceso de acondicionamiento para que satisfacer las necesidades de unos usuarios de la comunidad Universitaria, tanto internos como externos, lo cual constituye un programa de ejecución continua. De igual forma, en la norma internacional ISO/IEC 25012:2008 se encontró la definición de calidad como el “grado en que las características de los datos satisfacen las necesidades expresadas e implícitas cuando se utilizan en condiciones especificadas”.

Una vez revisado el marco normativo y de procedimientos de la Universidad Nacional, no se encontró acto administrativo o manual de procedimientos donde se indiquen las políticas de la Universidad en cuanto a condiciones de estandarización de la información en el momento del nacimiento del ciclo de vida del dato, formas de valorar la calidad de los datos y formas de hacer seguimiento y monitoreo a este proceso. Debido a lo anterior, se procedió a abordar estos tópicos dentro del documento.

Posterior a la fase I de este proyecto, se recomienda a) dar continuidad al monitoreo de indicadores de calidad de forma mensual y permanente e implementar las acciones de mejora sobre procedimientos que conducen a capturar información que no cumple con los indicadores de calidad; b) proponer indicadores de calidad para otras etapas del ciclo de vida del dato, entre las cuales se pueden citar almacenamiento, transformación y eliminación c) construir un diccionario de la información transversal a todas las dependencias de la Universidad que estandarice la definición de variables compartidas entre dependencias (lo que mitiga el riesgo de duplicidad y reprocesamiento de la información e impacta directamente sobre la reducción de tiempos de respuesta a solicitudes de información); d) identificar las reglas de negocio propias de cada dependencia que miden la calidad, ya no solo de la información en su momento de captura, sino de los procesos misionales, y construir perfiles de estos; e) cuantificar el riesgo de pérdida de la información presente en cada sistema; f) identificar lineamientos para la implementación de políticas asociadas a la seguridad de la información.

REFERENCIAS

- Ansari, D. (2017). *Data modeling 101: The what? Why? And when?* Inciter. <https://www.inciter.io/data-modeling-101-the-what-why-and-when/>
- Buenadicha, C et al. (2019) La Gestión Ética de los Datos. Por qué importa y cómo hacer un uso justo de los datos en un mundo digital. Banco Interamericano de Desarrollo. Disponible en: <https://publications.iadb.org/es/la-gestion-etica-de-los-datos>
- Burbank, D. (2020). *Best Practices in Metadata Management*. <http://www.dataversity.net/das-slides-best-practices-in-metadata-management/#>
- Chen, P. P.-S. (1976). The Entity-Relationship Model—toward a Unified View of Data. *ACM Trans. Database Syst.*, 1(1), 9–36. <https://doi.org/10.1145/320434.320440>
- Chisholm, M. (2001). The Foundations of Successful Reference Data Management. *TopQuadrant*, 1–16. https://www.topquadrant.com/docs/whitepapers/TopBraid_ReferenceDataManagementWhitepaper-3-18-15.pdf
- Congreso de Colombia. (2008). *Ley 1266 de 2008*. Diario Oficial 47.219 de diciembre 31 de 2008.
- Congreso de Colombia. (2012). *Ley Estatutaria 1581 de 2012*. Diario Oficial 48587 de octubre 18 de 2012.
- Congreso de Colombia. (2014a). *Decreto 886 de 2014*. Diario Oficial No. 49.150 de 13 de mayo de 2014.
- Congreso de Colombia. (2014b). *Ley 1712 de 2014*. Diario Oficial 49.084 de marzo 6 de 2014.
- Constitución Política de Colombia (2nd ed.). (1991). LEGIS.
- Corte Constitucional de Colombia. (2008). *Sentencia C-1011 de 2008*.
- DAMA Internacional. (2017). *DAMA-DMBOK. Guía del conocimiento para la gestión de datos*. Technics Publications.
- DataONE. (2016). *Tutorial on Data Management Lesson 7: Metadata*. <https://old.dataone.org/education-modules>
- Departamento Administrativo Nacional de Estadística [DANE]. (2005). *Guía para Diseño, Construcción e Interpretación de Indicadores*. Estrategia para el Fortalecimiento Estadístico Territorial.
- Departamento Administrativo Nacional de Estadística [DANE]. (2009). *Línea Base de Indicadores*. https://www.dane.gov.co/files/planificacion/fortalecimiento/cuadernillo/Linea_base_indicadores.pdf
- Evans, J. R., & Lindsay, W. M. (2015). *Administración y control de la calidad*. Cengage Learning Editores.
- Flórez, R. (2011). *La protección de la intimidad económica con relación al dato financiero*. Ediciones UNAULA.
- Gilliland, A. (2008). Setting the Stage. In M. Baca (Ed.), *Introduction to Metadata* (2nd ed., pp. 1–19). Getty Research Institute.

International Organization for Standardization - ISO. (2013). *Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes* (3rd ed.). https://standards.iso.org/ittf/PubliclyAvailableStandards/c050340_ISO_IEC_11179-3_2013.zip

International Organization for Standardization - ISO. (2015a). *ISO/IEC 2382:2015(en) Information technology – Vocabulary*. <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en>

International Organization for Standardization - ISO. (2015b). *ISO/IEC CD 11179-1 Information technology – Metadata registries (MDR) – Part 1: Framework*. <https://www.iso.org/standard/78914.html>.

International Organization for Standardization [ISO]. (2019). *ISO/IEC 25012:2008 Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model*. ICS - 35 -35.080. <https://www.iso.org/standard/35736.html>

International Organization for Standardization, & International Electrotechnical Commission. (2017). *ISO/IEC 38505-1:2017 Information technology – Governance of IT – Governance of data*.

International Organization for Standardization, & International Electrotechnical Commission. (s.f.) Fase 1. Auditoria Inicial ISO 27001 GAP Analysis. Disponible en: <https://normaiso27001.es/1-auditoria-inicial-iso-27001-gap-analysis/>

Kerri, S. (2020). *SIPOC Diagram*. ISIXSIGMA. <https://www.isixsigma.com/tools-templates/sipoc-copis/sipoc-diagram>

Kim, S., & Ji, Y. (2018). Gap analysis. The International Encyclopedia of Strategic Communication, 1-6.

Loshin, D. (2009). *Master Data Management*. Morgan Kaufmann Publishers.

Ministerio de Salud. (2021). *Ficha técnica del indicador IREM (Índice de Resistencia Epidemiológica Municipal)*. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/ficha-tecnica-inidcador-irem.pdf>

Ministerio de Tecnologías de la Información y las Comunicaciones. (2014). *G.INF.06 Guía Técnica de la información - Gobierno del dato*.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2016). *Guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia*. Gobierno de Colombia, MINTIC. https://herramientas.datos.gov.co/sites/default/files/A_guias_de_estandares_final_0.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (2019a). *Estándares de la industria*. https://mintic.gov.co/arquitecturati/630/propertyvalues-8158_descargable_2.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones. (2019b). *Guía para el uso y aprovechamiento de Datos Abiertos en Colombia*.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2019c). *Marco de interoperabilidad para Gobierno Digital*. https://mintic.gov.co/arquitecturati/630/articles-9375_recurso_4.pdf

Ministerio de Tecnologías de la Información y las Comunicaciones [MinTIC]. (2019). *G.INF.01 Guía del dominio de la información de TI*. https://www.mintic.gov.co/arquitecturati/630/articles-9253_recurso_pdf.pdf

Object Management Group. (2017). *Data Modeling Methodology and Notation*. Association for Retail Technology Standards.

Park, J. (2012). *Metadata Best Practices and Guidelines: Current Implementation and Future Trends* (1st ed.). Routledge.

Parlamento Europeo. (2016). *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo (27 de abril de 2016)*. Diario Oficial de la Unión Europea L 1119/1. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32016R0679>

Pervasive Software Inc. (2009). *A Metadata Management Strategy Checklist: What Metadata Really Matters?*

http://viewer.media.bitpipe.com/1013484794_935/1239305186_31/WhatMetaDataMemberMatters.pdf

Portal ISO 25000. (2005). *ISO/IEC 25000*. <https://iso25000.com/index.php/normas-iso-25000/iso-25010>

Presidencia de la República de Colombia. (2013). *Decreto 1377 de 2013*. Diario Oficial 48834 de junio 27 de 2013.

Queensland Goverment. (2019). *Metadata lifecycle management guideline*. Queensland Goverment Enterprise Arquitecture. <https://www.qgcio.qld.gov.au/documents/metadata-lifecycle-management-guideline>

Reed, M. (2007). Federal data architecture. Definitions and Structure. *The MIT Information Quality Industry Symposium*, 15. http://mitiq.mit.edu/IQIS/Documents/CDOIQS_200777/Papers/01_16_2B.pdf

Rodríguez, I. (2020). *Metadatos estadísticos para el aseguramiento de la calidad de los datos* [Universidad Nacional de Colombia, Sede Medellín]. <https://repositorio.unal.edu.co/bitstream/handle/unal/78418/527779.2020.pdf?sequence=3&isAllowed=y>

Rodríguez Rodríguez, A., & Bernal Gamboa, E. (2019). *Gestión de la información cuantitativa en las universidades. Pistas para su abordaje en la era de la sobreinformación*. Universidad Nacional de Colombia - Dirección Nacional de Planeación y Estadística.

Rummel, G. A., Ramias, A. J., & Rummel, R. A. (2010). *White Space Revisited. Creating value through process*. John Wiley & Sons.

Sansone, S.-A. (2017). *FAIR digital research assets: beyond the acronym*.

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2(3), 2–13.

Smith, A. M. (1997). How to Implement a Meta Data Strategy. *The Data Administration Newsletter*. <https://tdan.com/how-to-implement-a-meta-data-strategy/4144>

Sulaiman, S. (2012). Demystifying Metadata - A Practical Approach to Solving Critical Business Problems. In *Dataversity*. <https://www.youtube.com/watch?v=SKV1no3mPXc>

Superintendencia de Industria y Comercio. (2019). *Estudio de medidas de seguridad en el tratamiento de datos personales*. <https://www.sic.gov.co/sites/default/files/files/pdf/SIC Estudio de seguridad de datos personales 2019.pdf>

Taylor, D. (2021). Data Modelling: Conceptual, Logical, Physical Data Model Types. *Guru99*. <https://www.guru99.com/data-modelling-conceptual-logical.html>

The ODI (2022) Data Ethics Maturity Model.
Disponible en: <https://theodi.org/article/data-ethics-maturity-model-benchmarking-your-approach-to-data-ethics/>

U.S. Department of Defense. (2020). *Executive Summary: DoD Data Strategy. Unleashing Data to Advance the National Defense Strategy.* <https://media.defense.gov/2020/Oct/08/2002514180-1/-1/0/DOD-DATA-STRATEGY.PDF>

Universidad Nacional de Colombia. (2012). *Manual de Indicadores de proceso en el marco del Sistema de Mejor Gestión - SIMEGE.* http://www.cienciasagrarias.bogota.unal.edu.co/sites/default/files/IMGS/Guia_indicadores.pdf

Universidad Nacional de Colombia - Consejo Superior Universitario. (2019). *Acuerdo 316 de 2019 (Acta 12 del 3 de diciembre).* http://www.legal.unal.edu.co/rlunal/home/doc.jsp?d_i=94762

Universidad Nacional de Colombia - Dirección Nacional de Estrategia Digital. (2020). *La UNAL cuenta. Balance Social Vigencia 2020.* Universidad Nacional de Colombia. http://www.launalcuenta.unal.edu.co/fileadmin/user_upload/documentos/Informes/Nacional/2020/DNED_2020.pdf

Universidad Nacional de Colombia - Grupo de Gestión de Información. (2021). *Diagnóstico de Política de Gestión de la Información.*

Propuesta de Política de Gestión de la Información de la Universidad Nacional de Colombia, (2021) (testimony of Universidad Nacional de Colombia - Grupo de Gestión de Información).

Universidad Nacional de Colombia - Rectoría. (2018). *Resolución 11 de 2018 (04 de enero). “Por la cual se adopta el Compromiso Ético de la Universidad Nacional de Colombia y se establecen directrices para su implementación y seguimiento.”*

Universidad Nacional de Colombia - Rectoría. (2021). *Resolución 207 de 2021 (16 de abril). “Por la cual se establece la Política de Tratamiento de Datos Personales de la Universidad Nacional de Colombia y se deroga la Resolución No. 440 de 2019 de Rectoría.”*

Universidad Nacional de Colombia - SIGA. (2020). *Guia de Gestión del Cambio Universidad Nacional de Colombia.* <http://sigu.unal.edu.co/images/Modulos/Ova/U.GU.SIGA.002-Gua-Gestin-del-Cambio.pdf>

UNAL (2019) Guia para la administración de riesgos UNAL. Disponible en: http://sigu.unal.edu.co/images/contenido/componentes/Guia_para_la_administracion_de_riesgos_UN.pdf

Watt, A. (2014). Chapter 5 Data Modelling. In A. Watt & N. Eng (Eds.), *Database Design* (2nd ed.). BCcampus. <https://opentextbc.ca/dbdesign01/>

Wilkinson, M., Dumontier, M., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3.

Yourdon, E., & DeMarco, T. (1979). DataFlow System. In *Structured Analysis and System Specification* (2nd ed., pp. 409–424).

Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. *El Profesional de La Información*, 28(1).

GLOSARIO

A

ACTIVO: aquello que tiene valor para una organización, estos activos dentro de una empresa se dividen en activos tangibles e intangibles.

ACTIVO TANGIBLE: activos materiales que contienen información, y sobre los que se tomarán medidas preventivas para protegerlos principalmente de riesgos físicos.

ACTIVO INTANGIBLE: soportan la información dentro de un activo material, y pueden inutilizar la información, pese a que el activo físico no haya sufrido daño alguno.

ACOPLAMIENTO DE LA APLICACIÓN: describe el grado en que dos sistemas se entrelazan. Dos sistemas que están estrechamente acoplados suelen tener una interfaz síncrona, donde un sistema espera una respuesta del otro, por lo cual la disponibilidad de ambos se encuentra ligada. En general se procura manejar bajos niveles de acoplamiento donde se permite la implementación de servicios de datos, APIs (*Application Programming Interface*) o colas de mensajes.

ADM: método de desarrollo de Arquitectura definido por TOGAF, este método es ajustado y personalizado según los intereses (*concerns*) o necesidades propias de la organización, el método es un modelo iterativo que establece un conjunto de buenas prácticas para llevar a la organización a una alineación tecnológica de los procesos y objetivos estratégicos.

ADMINISTRACIÓN DE DATO MAESTRO: es el proceso mediante el cual, se define y gestiona la manera en que la organización

crea, mantiene y usa datos maestros coherentes, completos y precisos en todas las partes de la institución (usuarios, ciudadanos, aplicaciones, bodegas de datos y procesos).

ADMINISTRADOR DEL DATO: es el encargado de orquestar las necesidades y las acciones técnicas, así como mediar en los posibles conflictos que puedan surgir entre roles consumidores y productores. Para esto, crea estándares y buenas prácticas, habilita metadatos técnicos, operacionales y funcionales, audita la calidad de los datos y las medidas de gobernanza, y define políticas de respaldo de la información (*backups*), seguridad, flujos de datos, etc.

ARCHIVADO: movimiento de datos con baja frecuencia de uso hacia unidades de almacenamiento menos costosas.

B

BASE DE DATOS: es una colección organizada de la información estructurada, o datos, típicamente almacenados electrónicamente en un sistema de computadora.

BUS DE SERVICIO EMPRESARIAL (ESB - Enterprise Service Bus): es un sistema que actúa como intermediario entre sistemas, pasando mensajes entre ellos. Las aplicaciones pueden enviar y/o recibir mensajes y/o archivos utilizando este medio, los cuales se encuentran encapsulados de otros procesos existentes en dicho Bus.

C

CAPTURA DE CAMBIOS EN DATOS: método de optimización de cargues de datos filtrando solamente aquellos que han llegado recientemente a las fuentes, evitando así realizar un análisis completo de toda la tabla.

CARDINALIDAD Y ORDINALIDAD: la cardinalidad captura el número máximo de elementos instanciados de una entidad que pueden participar con las otras instancias de la otra entidad, en cuanto a la ordinalidad es el número mínimo de elementos de una entidad que deben relacionarse con las instancias de la otra entidad.

CONSUMIDORES: son aquellos que se benefician directamente de los productos primarios generados por las actividades de gestión de datos.

CONOCIMIENTO: proceso de apropiación de la información resultado del procesamiento de datos. Así como la información se deriva de los datos, el conocimiento se deriva de la información. El conocimiento tiende a ser progresivo a través del tiempo, por ello permite a los decisores y ejecutores aprender acerca de dinámicas y comportamientos soportados en evidencias reales y no de forma intuitiva.

CRUD: cuatro operaciones básicas de almacenamiento persistente de datos, Crear, Leer, Actualizar y Eliminar.

CUSTODIO DEL DATO: responsable por la existencia de unos datos en la función del sector público, por lo que cualquier acción correctiva o nuevo requisito que precise su información, debe ser consensuado con él. Igualmente es el responsable de que el dato esté disponible y con atributos de calidad. Adicionalmente, este actúa como patrocinador de calidad y control sobre estos datos, y es considerado maestro en conocimiento sobre el dato que es custodio.

D

DATA LAKE: es un entorno donde una gran cantidad de datos de varios tipos y estructuras

pueden ser ingeridos, almacenados, evaluados y analizados.

DATA WAREHOUSE: es un sistema que agrega y combina información de diferentes fuentes en un almacén de datos único y centralizado; consistente para respaldar el análisis empresarial, la minería de datos, inteligencia artificial (IA) y *Machine Learning*. *Data warehouse* permite a una organización o empresa ejecutar análisis potentes en grandes volúmenes (petabytes) de datos históricos de formas que una base de datos estándar simplemente no puede.

DATO: es una representación simbólica cuantitativa o cualitativa de una característica particular de la realidad que se obtiene a través de una captura o transferencia. Los datos no proveen juicios de valor o interpretaciones dado que necesitan un contexto para adquirir su significado y uso, por ello, no son orientativos para la acción.

DATOS ANALÍTICOS: describen el rendimiento de la organización.

DATOS TRANSACCIONALES: describen el estado de las operaciones de la organización.

DIAGRAMA SIPOC: representación del mapa de proceso de negocio a través de la identificación de proveedores (*Suppliers*), entradas (*Inputs*), procesos (*Processes*), productos (*Outputs*) y consumidores (*Consumers*). Es también conocido como diagrama COPIS.

DISEÑO DE ARQUITECTURA DE DATOS: representación a nivel global de la ubicación de los datos en la organización, la cual sirve como plantilla para la elaboración del modelo de datos empresarial.

DOMINIO: es un conjunto con todos los posibles valores que se pueden asignar a un atributo de la misma índole, un ejemplo de ello es la fecha,

todos los atributos que relacionan fechas pueden tomar diferentes formatos de fecha como aaaa/mm/dd, o dd/mm/aaaa por nombrar unos, estos posibles valores hacen parte del dominio de fecha. Son las dimensiones que agrupan y organizan los lineamientos del Modelo.

DOMINIO DE ARQUITECTURA DE APLICACIONES EMPRESARIAL: descripción de estructura y funcionalidad de las aplicaciones en la empresa. Proporciona un plan para las aplicaciones individuales que se implementarán, sus interacciones y sus relaciones con los procesos institucionales básicos de la organización.

DOMINIO DE ARQUITECTURA DE DATOS EMPRESARIAL: descripción sobre cómo los datos deben estar organizados y administrados. Define los términos y diseños estandarizados para los elementos que son importantes para la organización.

DOMINIO DE ARQUITECTURA DE NEGOCIO EMPRESARIAL: identificación de cómo una empresa crea valor para sus clientes y otros interesados; define la estrategia empresarial, la gobernanza, la organización y los procesos empresariales clave.

DOMINIO DE ARQUITECTURA EMPRESARIAL: conjunto de dominios de conocimiento que se tienen en cuenta al interior de una empresa u organización.

El marco TOGAF tiene cuatro dominios de arquitectura principales: negocios, datos, aplicaciones y tecnología. También se pueden considerar otros dominios (por ejemplo, seguridad).

DOMINIO DE ARQUITECTURA TECNOLÓGICA EMPRESARIAL: descripción de la tecnología física necesaria para habilitar el funcionamiento de los sistemas y entregar valor; además, describe las capacidades lógicas de software y hardware

Código: U.MD.17.001.001

Versión: 0.0

necesarias para apoyar el despliegue de servicios empresariales, de datos y de aplicaciones; esto incluye infraestructura de TI, middleware, redes, comunicaciones, procesamiento, normas, etc.

DUEÑO DEL DATO: entidad que puede autorizar o denegar el acceso a determinados datos, y es responsable de su veracidad, integridad y actualidad.

DATO: se argumenta que los datos deben definirse como información sobre las propiedades de las unidades de análisis. Epistemológicamente es importante establecer que lo que alguien considera datos no tiene por qué ser datos de otra persona. Este artículo considera la naturaleza de los datos y los "macrodatos" y la relación entre datos, información, conocimiento y documentos. Es común que todos estos conceptos se refieran a fenómenos producidos en contextos específicos para fines específicos y pueden estar representados en documentos, incluso como representaciones en bases de datos. En ese proceso, se sacan de sus contextos originales y se colocan en otros nuevos y, por lo tanto, los datos pierden parte o todo su significado debido al principio del holismo semántico. Parte de este significado perdido debe restablecerse en las bases de datos y las representaciones de datos / documentos no pueden entenderse como una actividad neutral, sino como una actividad que apoya el objetivo general implícito en el establecimiento de la base de datos. La utilización de (grandes) datos (como es el caso de la utilización de información, conocimientos y documentos) exige en primer lugar la identificación de los potenciales de estos datos para fines relevantes. Un aspecto importante sobre los macrodatos es que a menudo son rastros involuntarios que dejamos durante todo tipo de actividades. Por lo tanto, su potencial para

Página 165 de 171

informar a alguien sobre algo es menos directo en comparación con los datos que se han producido intencionalmente como, por ejemplo, bases de datos científicas.

DATOS ABIERTOS: son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen

DATOS ESTRUCTURADOS EN LAM: generalmente se almacenan en bases de datos en las que todos los pares clave/valor tienen identificadores, relaciones claras y siguen un modelo de datos explícito (Schöch, 2013).

DATOS SEMIESTRUCTURADOS EN LAM: comprenden las secciones no estructuradas dentro de las descripciones de metadatos. documentación de archivo no llevado en Descripción de archivo codificado (EAD) u otras ayudas digitales para la búsqueda, trabajos intelectuales codificados siguiendo las pautas de la Iniciativa de codificación de texto (TEI) (cuerpo del texto principal excluyendo el encabezado), valor agregado o etiquetado recursos que existen en todo tipo de formatos, y porciones no estructuradas de conjuntos de datos (Zeng, 2019).

DATOS NO ESTRUCTURADOS EN LAM: son caracterizados como "todo lo demás". En LAM se pueden encontrar en documentos y otros objetos que contienen información. (textual o no textual, digitalizado o no digitalizado), en todo tipo de formatos.

Están disponibles en la mayor cantidad en comparación con los datos estructurados, tienen la mayor diversidad en tipo, naturaleza y calidad, y son los más difíciles de procesar. (Zeng, 2019, pp. 3-4).

DOMINIO: es un área particular de conocimiento o campo de interés en estudio.

ESQUEMA RELACIONAL: es aquel que busca tener una expresión exacta de datos de negocio y tener un hecho en un sólo lugar (eliminación de la redundancia). el esquema relacional es ideal para el diseño de sistemas operativos, que requieren introducir información rápidamente y que se almacena con precisión.

ETL (Extraer, Transformar, Cargar): actividades básicas realizadas en los procesos de integración de datos, extrayendo de una o más fuentes de datos, para luego realizar transformaciones sobre los mismos (que pueden incluir consolidación, validación, estandarización) para finalmente cargar dichos datos sobre una bodega, donde se dejan disponibles para consultas.

F

FORMATO DE MENSAJE EMPRESARIAL / MODELO CANÓNICO: modelo unificado, utilizado por una organización o grupo de intercambio de datos, que hace homogéneo el formato en el que se compartirán los datos. Los datos se transforman de o hacia un sistema de envío o de recepción, con base en un formato de mensaje común, o empresarial, para toda la organización.

FUENTE DE INFORMACIÓN: son aquellos recursos que permiten satisfacer una necesidad informativa. Las fuentes de la información se pueden clasificar acorde al tipo de la información que manejan, si esta es estructurada o no estructurada. La información estructurada se refiere a los datos que están perfectamente definidos y sujetos a un formato muy concreto, en una base de datos son campos que ya tienen una definición específica, y la información no estructurada son aquellos datos que no están en una base de datos o están

contenidos en algún otro tipo de estructura de datos, como repositorios documentales, fotos o correos electrónicos.

G

GESTOR DEL DATO: es el encargado del direccionamiento estratégico para gobernar el valor y uso del dato, y de definir las reglas de consolidación y coordinación con los administradores de unos datos en sus respectivos dominios. A veces, se encarga de la corrección manual de la información en caso de que las reglas generales no sean válidas para algún caso concreto.

H

HERENCIA: proceso mediante el cual una subentidad hereda los atributos de una entidad maestra.

I

INFORMACIÓN: es un conjunto organizado de datos que tienen un significado, propósito, relevancia y contexto. La información es el producto de la contextualización de los datos, siendo así enfocados hacia un fin específico.

INTEGRACIÓN: está relacionada con el intercambio unidireccional de datos a partir de una o varias fuentes (extracción) hacia uno o varios destinos realizando en el camino una transformación de esos datos (que puede ser estandarización, consolidación, verificación de métricas de calidad, generación de registros con errores, entre otras) y cargar en un destino de la información apropiado para realizar tanto reportes estandarizados como consultas analíticas. Por ende, la integración está usualmente relacionada con la construcción de bodegas de datos (*Datawarehouse - DWH*) y el ámbito de aplicación es sobre conjuntos

Código: U.MD.17.001.001

Versión: 0.0

masivos de datos (Ej: cargue de admitidos en el SIA o en una bodega de datos de integración).

INTEROPERABILIDAD: es la 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 digitales a ciudadanos, empresas y a otras entidades, mediante el intercambio de datos entre sus sistemas TIC. Está relacionada con el intercambio bidireccional de datos de una o varias fuentes hacia uno o varios destinos donde se usa un lenguaje de comunicación previamente definido y usando como mesa de trabajo un entorno común de acceso, donde se publica y se consume cierta información específica definida por las áreas de negocio. Por ende, está directamente relacionado con la construcción de un Bus Empresarial de Servicios (*Enterprise Service Bus - ESB*) orientado a la interoperabilidad y el ámbito de aplicación es sobre registros o transacciones específicas de datos.

INTEROPERABILIDAD EXTERNA: procesos de interoperabilidad con procesos y/o sistemas externos a la Universidad. Ejemplo: Interoperabilidad con la Registraduría Nacional del Estado Civil para la validación de documentos de identidad.

INTEROPERABILIDAD INTERNA: procesos de interoperabilidad dados entre procesos y/o sistemas al interior de la Universidad, por lo cual puede usar un lenguaje común definido también internamente. Ejemplo: Interoperabilidad SARA-QUIPU.

L

LATENCIA: diferencia de tiempo entre la generación de los datos en el sistema de origen y cuando están disponibles para su uso en el **Página 167 de 171**

sistema de destino. Puede ser alta (procesamiento por lotes), baja (impulsada por **EVENTOS**) O MUY BAJA (**TIEMPO REAL - SÍNCRONA**).

LÍDER FUNCIONAL: funcionario perteneciente al área de negocio, responsable del manejo funcional de cada sistema de la información y/o aplicación.

LÍDER TÉCNICO: funcionario perteneciente a la Oficina de Tecnologías de la Información y las Comunicaciones, que tiene el conocimiento técnico necesario para atender los requerimientos realizados por un área funcional para un determinado sistema de la información o aplicación.

LLAVE: en un modelado Entidad-Relación se pueden identificar dos tipos de llaves: Llave primaria: es un atributo que identifica de forma única a una instancia de una entidad. Llave foránea: es utilizada en esquemas de modelo de datos lógico. Es un atributo creado de manera implícita, cuando existe una relación de dependencia o agregación entre dos entidades, y cuyo valor corresponde al de una llave primaria de alguna de las entidades involucradas en la relación.

LOTE: Conjunto definido de datos que se mueven en grupo, como parte de un proceso batch asíncrono.

M

MAPA DE RUTA: es un conjunto estructurado de acciones que define la manera de lograr los objetivos fijados en una estrategia. Un mapa de ruta está expresado en términos de programas o proyectos, que son agrupadores de las acciones, y tiene asociados estimaciones de tiempo, costos y riesgos.

MARCO DE INTEROPERABILIDAD: es la estructura de trabajo común donde se alinean los

conceptos y criterios que guían el intercambio de información. Define el conjunto de principios, recomendaciones y directrices que orientan los esfuerzos políticos, legales, organizacionales, semánticos y técnicos de las entidades, con el fin de facilitar el intercambio seguro y eficiente de información.

MATRIZ CRUD: identifica qué tratamiento realiza cada sistema sobre cada dato, utilizando operaciones básicas como: crear, leer, actualizar y eliminar. Con base en este tratamiento, se pueden identificar posibles duplicidades e inconsistencias, así como realizar validaciones cruzadas.

METADATO: son datos sobre los datos. Los metadatos articulan un contexto para determinados objetos de interés (recursos), en forma de descripción de recursos.

MODELADO DE DATOS: es el proceso de descubrir, analizar y definir el alcance de los requerimientos de los datos, con el objetivo de documentar, representar y comunicar su estructura y acople en conjunto.

MODELO CONCEPTUAL (CDM): es aquel donde se capturan solo las entidades básicas y fundamentales para el ámbito y la función del negocio, dando una descripción y las relaciones existentes entre ellas.

MODELO DE DATOS: Abstracción mediante la cual se estructura la información, mostrando relaciones entre entidades, las limitaciones que contiene y cómo se almacenan los datos, también muestra la manera de acceder a ellos. Existen de tres tipos dependiendo el nivel de abstracción, siendo el de mayor nivel el conceptual, seguido del lógico y por último el físico.

MODELO DE DATOS EMPRESARIAL (ENTERPRISE DATA MODEL - EDM): es un modelo de datos conceptual o lógico holístico, a nivel

de empresa, independiente de la implementación que proporciona una visión común y consistente de los datos en toda la empresa. Un EDM incluye entidades de datos empresariales clave, sus relaciones, reglas de negocio y algunos atributos críticos. Establece la base para todos los datos y proyectos relacionados con datos.

MODELO DE INTERACCIÓN: diferentes maneras de establecer conexiones entre sistemas. Pueden ser a) directas punto a punto, o b) modelo de rueda (concentrador de radios) que aplica para los buses de servicio empresarial (ESB - Enterprise Service Bus) o también c) Publicar-suscribir donde un elemento de la red ofrece unos servicios de datos a los cuales otros elementos de la red se suscriben y reciben actualizaciones en cuanto son publicadas por el origen.

MODELO FÍSICO (PDM): representa una solución técnica detallada, generalmente utilizar como punto de partida el modelo lógico, el cual es adaptado a un conjunto de hardware, software y herramientas de red, para así dar solución al problema presentado. Su construcción se basa en la utilización de una tecnología en particular.

MODELO LÓGICO (LDM): es una representación detallada de las entidades, haciendo énfasis en los requerimientos de datos, los modelos lógicos son independientes de cualquier tecnología. A menudo, es una extensión de un modelo de datos conceptual.

O

ORQUESTACIÓN: se refiere al cómo se organizan y ejecutan múltiples procesos en un sistema. Todos los sistemas que manejan mensajes o paquetes de datos deben ser capaces de manejar el orden de ejecución de dichos procesos, con el fin de preservar la consistencia y la continuidad.

Código: U.MD.17.001.001

Versión: 0.0

P

PARTICIPANTES: son las personas que realizan, administran el desempeño o aprueban las actividades en el Área de Conocimiento.

PRINCIPIOS FAIR: directrices para mejorar la localización, accesibilidad, interoperabilidad y reutilización de los datos.

PROCESOS: corresponde a la definición del ciclo de vida de la información, desde su entrada o creación, a través de cualquiera de los sistemas, hasta su salida o retorno a los sistemas fuente de información; así mismo, tiene en cuenta el seguimiento de calidad de los datos y su variación a lo largo de su ciclo de vida y aseguramiento del cumplimiento de las necesidades de cada proceso.

PROVEEDORES: son las personas responsables de proporcionar o permitir el acceso a los insumos (entradas) para las actividades.

PUNTO DE INICIO: es un tipo de relación, que mediante un proceso crea, mantiene o actualiza los datos de una entidad.

R

RELACIÓN: es una asociación entre dos o más entidades, en ella se capturan las interacciones de alto nivel en el modelo conceptual y la interacción detallada entre las entidades en el modelo lógico. Las relaciones se identifican mediante verbos en presente simple.

REPLICACIÓN: copia de datos sobre diferentes servidores bien sea físicos o lógicos con objeto de mejorar la disponibilidad del servicio. Se requiere un proceso posterior de sincronización para mantener la consistencia de las diferentes copias.

REPRESENTACIÓN DE ENTIDAD: en un modelado de datos, una entidad maestra es un objeto en el cual una organización recaba información. Las entidades se identifican mediante sustantivos los cuales deben estar en su forma singular a menos de que el concepto mismo sea plural.

ROLES: conjunto de responsabilidades y actividades asignadas a una persona o grupo de personas para apoyar la adopción y aplicación de un Marco o modelo de Referencia.

S

SISTEMA DE INFORMACIÓN: conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar los procesos de toma de decisiones y de control en una organización.

T

TOMA DE DECISIONES: Es un proceso mediante el cual se usa el conocimiento para discernir entre diferentes alternativas con el propósito de resolver un problema actual o potencial en contextos diversos. La toma de decisiones es un proceso que puede ajustarse a través del tiempo de acuerdo con los cambios que surjan en los datos, la información, o el conocimiento.

Elaboró:	Sergio Iván Martínez Clúster de Gestión de la Información	Revisó:	Klinger Blanco	Aprobó:	Johan Sebastián Eslava
Cargo:	Profesional de apoyo de la Dirección Nacional de Estrategia Digital Clúster de Gestión de la Información	Cargo:	Profesional especializado	Cargo:	Director Nacional de Estrategia Digital
Fecha:	15/10/2023	Fecha:	15/10/2023	Fecha:	12/12/2023