



El futuro digital
es de todos

MinTIC

PNID

Anexo 1.2

Ciclo de vida del dato

MinTIC, DNP, DAPRE
Diciembre 2021

- Carmen Ligia Valderrama Rojas - Ministra de Tecnologías de la Información y las Comunicaciones
- Iván Mauricio Durán Pabón - Viceministro de Transformación Digital
- Ingrid Tatiana Montealegre Arboleda – Directora de Gobierno Digital
- Gersson Jair Castillo Daza - Subdirector de Estándares y Arquitectura de TI

- Jairo Alberto Riascos Muñoz – Líder de equipo de Estándares y Arquitectura
- Marco Emilio Sanchez Acevedo - Coordinación de Política
- Thiago Grijo Dal Toe - Equipo de Datos Abiertos
- Carlos Julio León Caicedo – Equipo de Datos Abiertos
- Gina Paola Mahecha Ortiz- Equipo de Estándares y Arquitectura
- Julio Cesar Anaya - Equipo de Estándares y Arquitectura
- Luisa Fernanda Medina Martínez – Líder de equipo de Datos Abiertos
- José Ricardo Aponte - Equipo Servicios Ciudadanos Digitales
- Alexander Alfonso Pérez – Lenguaje común de intercambio

Departamento Nacional de Planeación

- Alejandra Botero Barco - Directora General del Departamento Nacional de Planeación
- Daniel Gómez Gaviria - Subdirector General Sectorial
- Viviana Rocío Vanegas Barrero - Directora de Desarrollo Digital
- Javier Lesmes – Subdirector de Prospectiva Digital
- Diana Ramírez Roa – Asesora - Dirección de Desarrollo Digital
- Liliana Fernández Gómez – Asesora - Dirección de Desarrollo Digital
- Agustín Jiménez Ospina – Asesor - Dirección de Desarrollo Digital
- Nicolás Enrique Chibuque Pérez – Asesor - Dirección de Desarrollo Digital

Departamento Administrativo de la Presidencia de la República

- Víctor Manuel Muñoz Rodríguez - Director de Departamento Administrativo de la Presidencia de la República
- María Lucia Villalba Gómez – Consejera Presidencial para Transformación Digital y Gestión y Cumplimiento
- Pablo Gómez Ayerbe – Asesor - Consejería Presidencial para Transformación Digital y Gestión y Cumplimiento

Centro para la Cuarta Revolución Industrial de Colombia - C4IR

- Erez Zaionce – Director Centro para la Cuarta Revolución Industrial
- Eduardo Gómez Restrepo – Líder Senior de Política de Datos

Versión	Observaciones
Versión 1 Diciembre 2021	Plan Nacional de Infraestructura de Datos – Documento técnico y hoja de ruta

Comentarios, sugerencias o correcciones pueden ser enviadas al correo electrónico: gobiernodigital@mintic.gov.co.

Anexo 1.2 Plan Nacional de Infraestructura de Datos – Ciclo de vida del dato



- Este documento de la Dirección de Gobierno Digital se encuentra bajo una Licencia Creative Commons Atribución 4.0 Internacional.

Tabla de contenido

Ciclo de vida del dato	4
------------------------------	---

1. Ciclo de vida del dato

- La gestión del ciclo de vida del dato es un proceso indispensable para alcanzar el aprovechamiento de los datos dispuestos en la infraestructura de datos del Estado. El ciclo de vida de los datos está integrado por diferentes fases como se muestra en la figura 1.



Figura 1. Fases del ciclo de vida del dato
Fuente: Adaptación Banco Mundial (Data for Better lives)

En la Tabla 1 se describe cada una de las fases del ciclo de vida:

Tabla 1. Descripción del ciclo de vida de los datos

Fase ciclo de vida	Descripción
Crear y obtener	Los datos pueden provenir de diferentes fuentes, estructuradas, no estructuradas o fuentes secundarias como las redes sociales; asimismo también son diversos los formatos en que pueden venir (Ej. pdf, jpg, docx, xml, txt, json, csv, png, entre muchos otros). Los datos suelen ser creados por las organizaciones usualmente de las siguientes formas:

	<ul style="list-style-type: none"> • Adquisición de datos: adquisición de datos ya existentes que se han producido fuera de la organización. • Entrada de datos: entrada manual de nuevos datos por parte del personal dentro de la organización. • Captura de datos: captura de datos generados por dispositivos utilizados en varios procesos de la organización. <p><i>El modelado y diseño de datos es de suma importancia en esta fase, pues determina que información es la que se determina utilizar.</i></p>
Procesar	<p>La etapa de procesamiento incluye las siguientes tareas:</p> <ul style="list-style-type: none"> • Limpieza de datos: en la que un conjunto de datos se limpia y se transforma de su forma sin procesar a algo más accesible y utilizable. Esto también se conoce corrección de datos. • Compresión de datos: en la que los datos se transforman en un formato que se puede almacenar de manera más eficiente. • Cifrado de datos: en el que los datos se traducen a otra forma de código para protegerlos de problemas de privacidad. • Calidad de los datos: en los proceso de obtención y procesamiento de datos, donde principalmente se aplican las políticas y controles, establecidas en el diseño, con el propósito de determinar cuáles datos son utilizables. <p>Además de las tareas de esta etapa mencionadas anteriormente existen muchas otras. incluso el simple hecho de tomar un formulario impreso y digitalizarlo puede considerarse una forma de procesamiento de datos.</p>
Almacenar	<p><i>El almacenamiento y operaciones</i> incluyen el diseño, la implementación y el soporte de los datos almacenados para maximizar su valor, utilizando tanto formatos como repositorios que busquen el equilibrio entre disponibilidad y coste de almacenamiento, según los distintos escenarios de consulta, por ejemplo, diferenciando datos de alta demanda o disponibilidad de otros que no se consultan frecuentemente.</p> <p>Comúnmente el almacenamiento se realiza mediante la creación de bases de datos o conjuntos de datos. Estos conjuntos de datos pueden almacenarse en la nube, en servidores <i>on premise</i> o utilizando otras formas de almacenamiento físico como discos duros magnéticos o de estado sólido, memorias o cintas magnéticas, entre otros.</p>
Transferir y compartir	<p>A medida que las organizaciones requieren de datos de fuentes secundarias, generados por otros actores del ecosistema de datos, útiles para la toma de decisiones, la planificación, la optimización de operaciones, entre otros, existe una mayor presión para compartir también datos generados. Los datos compartidos pueden ayudar a mejorar las decisiones, ya que las organizaciones pueden obtener una vista más completa de los impactos que sus decisiones han tenido en base a las contribuciones de nuevos conjuntos de datos de una variedad más amplia de fuentes, tanto internas como externas. Esta capacidad de compartición de datos debe estar soportada en una capa de interoperabilidad y haciendo uso de un estándar de lenguaje común de intercambio.</p>

<p>Analizar y usar</p>	<p>Durante la fase de uso del ciclo de vida de los datos, los datos se utilizan para respaldar las actividades de la organización. Los datos se pueden ver, procesar, modificar y guardar. Se debe mantener un registro de auditoría para todos los datos críticos, con el propósito de garantizar que todas las modificaciones que se realicen a los datos sean completamente rastreables y auditables. Los datos también pueden estar disponibles para exponer a otros actores del ecosistema de datos que se encuentren fuera de la organización.</p> <p>El análisis de datos se refiere a procesos que intentan obtener información significativa a partir de datos sin procesar. Los analistas y científicos de datos utilizan diferentes herramientas y estrategias para realizar estos análisis. Algunos de los métodos más utilizados incluyen modelado estadístico, algoritmos, inteligencia artificial, minería de datos y aprendizaje automático.</p>
<p>Archivar y preservar</p>	<p>El archivado de datos hace referencia a la copia de datos en un entorno donde se almacenan en caso de que se necesiten nuevamente en un ambiente de producción, así mismo también incluye la eliminación de estos datos de todos los entornos de producción activos.</p> <p>Un archivo de datos es simplemente un lugar donde se almacenan los datos, pero donde no se realiza mantenimiento o uso general. Si es necesario, los datos se pueden restaurar a un entorno en el que se puedan utilizar.</p> <p>Dado que el volumen de datos archivados crece inevitablemente, si bien es posible que desee guardar todos los datos de manera indefinida, los costos de almacenamiento pueden incentivar la destrucción de los datos que ya no se requieren. Por otro lado, es posible que se disponga de datos de uso limitado a una ventana de tiempo o a hasta /durante la ocurrencia de un suceso. La destrucción o depuración de datos es la eliminación de cada copia de un elemento de datos de una organización.</p>
<p>Reutilizar</p>	<p>Reutilizar significa usar datos que originalmente fueron recopilados para otro fin o propósito. La reutilización de datos también se puede llamar análisis secundario.</p>

Fuente: elaboración propia