TIC

PROYECTO IA PARA EL ESTADO

E8: Documento estrategia datos abiertos Colombia

Universidad de Cartagena

2026

Universidad de Cartagena
Fundada en 1827

Créditos

Equipo MINTIC

Dirección de Gobierno Digital, MinTIC
Jaime Emiro Criales Henao

Subdirectora de Fortalecimiento de Capacidades Públicas Digitales
Luisa Fernanda Medina Martínez

Subdirectora de Estándares de Arquitectura
Lucy Palacios Valoyes

Apoyo Supervisión
Lina María Lara Sanabria
Jennifer Vivian Prada Urbina
Carlos Julio León Caicedo

Equipo Universidad de Cartagena

Coordinador
Juan Carlos Vergara Schmalbach

Equipo Componente 1
Francisco Javier Maza Ávila
José Romano Burgos
Harold Lora Guzman
José Morelos Gómez
Rosario Blanco Hernández
Álvaro Andres Escobar
Lina María Velandia Rios
Edwin Mora Cepeda

Equipo Componente 2
Henry Palacio Brochero
Eyvis Lucia Marrugo Batista
Victor Valdelamar García
Plinio Puello Marrugo
Danilo Andrés Castrillón
Diego Osorio Gutierres
José David Pérez
Andrés Felipe Chavez
Juan David Gallego

Equipo Componente 3
Oscar Zuñiga Fernández
Silvia Angarita Jara
Emiro José Florez Mendez
William Diaz Quintana
María Fernanda Martínez
Juan Camilo Montaño
Fabián Andrés Flórez Rodríguez
Jhonatan Leonardo Ramos

Equipo Transversal
Hector Morantes Díaz
Ana Banquez Maturana
Natalia Torres Luenga
Farid Baquedo Fernandez
Juan Alberto Cardenas

Supervisión
Katherine Porto Martínez

Contenido

1. Resumen 4
2. Introducción y visión estratégica 5
2.1. Propósito de la estrategia 5
2.2. Alcance y público objetivo 5
3. Marco normativo y de política pública 8
4. Contexto 11
4.1. Tendencias internacionales 11
4.2. Diagnóstico del catálogo nacional de datos abiertos para IA 12
4.2.1 Concentración y madurez de la reutilización 14
4.2.2 Formatos, acceso y licenciamiento 15
4.3. Retos actuales 16
4.4. Síntesis DOFA articulada a los ejes estratégicos 17
5. Modelo estratégico de datos abiertos para inteligencia artificial 20
5.1. Principios 20
5.2. Objetivos y ejes 20
5.3. Ecosistema de actores 21
6. Datos abiertos para inteligencia artificial 23
6.1. Datos listos para inteligencia artificial y criterios mínimos 23
6.2. Alto valor, datos críticos y datos estratégicos 23
6.3. Muestra priorizada de conjuntos con aptitud para IA 24
6.4. Riesgos 26
6.5. Aspectos éticos 28
6.6. Casos nacionales e internacionales 28
7. Fortalecimiento de metadatos y catálogos 30
7.1. El esquema de metadatos actual de datos.gov.co 30
7.2. Campos mínimos de metadatos para IA 30
7.3. Vocabularios, taxonomías e interoperabilidad semántica 32
7.4. Automatización del levantamiento de metadatos 32
7.5. El catálogo orientado a IA: sección del portal datos.gov.co 32
8. Desarrollo de capacidades 34
8.1. Ruta de formación para entidades 34
8.2. Ruta de formación para desarrolladores 34
8.3. Certificación, comunidad e indicadores 36
9. Guía para la implementación, desarrollo y uso de sistemas de inteligencia artificial en entidades públicas de colombia 37
9.1. Paso 1. Identificar los datos de alto valor 38
9.2. Paso 2. Priorizar 38
9.3. Paso 3. Asegurar la calidad con la autoevaluación de aptitud para IA 38
9.4. Paso 4. Documentar los metadatos 40
9.5. Paso 5. Publicar 40

9.6. Paso 6. Actualizar 41
9.7. Paso 7. Gobernar 41
9.8. Paso 8. Hacer seguimiento 42
10. Guía para reutilizadores 43
10.1. La metodología CRISP-ML como ruta de trabajo 43
10.2. Cómo encontrar los datos 45
10.3. Acceso mediante interfaces de programación 45
10.4. Casos de uso como material de aprendizaje 45
10.5. Desarrollo responsable de sistemas de inteligencia artificial 46
10.6. Buenas prácticas del reutilizador 47
Referencias 48
Anexo A. Evidencias 51
A1. Vigencia de los datos 51
A2. Legibilidad por máquina 51
A3. Consistencia de metadatos 52
A4. Uniformidad del licenciamiento 52
A5. Foco analítico del catálogo 53
A6. Curaduría de activos comunitarios 53
A7. Ampliación de la reutilización 53
Anexo B. Matriz de trazabilidad 55
B.1. Trazabilidad contractual con el anexo técnico 55
B.2. Trazabilidad con el CONPES 4144 de 2025 56
B.3. Trazabilidad con el marco normativo 58
B.4. Trazabilidad interna entre ejes, instrumentos e indicadores 59
Anexo C. Machine Learning Canvas 60
C.1. Función y momento de aplicación 60
C.2. Estructura del lienzo adaptado a datos abiertos 60
C.3. Criterios de decisión sobre la viabilidad 62
Anexo D. Ficha de caso de uso de inteligencia artificial 64
D.1. Función de la ficha y ruta de publicación 64
D.2. Estructura de la ficha 64
D.2. Criterios de aceptación editorial 67
Anexo E. Matriz de indicadores 69
E.1. Estructura, convenciones y universo de medición 69
E.2. Matriz consolidada de indicadores 69
E.3. Fichas técnicas 70

1. Resumen

El presente documento formula la Estrategia de Datos Abiertos de Colombia orientada al desarrollo y la adopción de sistemas de inteligencia artificial en el sector público, en el marco del convenio entre el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) y la Universidad de Cartagena. La estrategia parte de un principio operativo verificable: la calidad, la disponibilidad y la aptitud técnica de los datos determinan el límite superior del desempeño de cualquier modelo que el Estado construya, contrate o supervise.

El diagnóstico realizado sobre el Portal Nacional de Datos Abiertos (datos.gov.co), consultado el 1 de julio de 2026, registra 8.422 conjuntos de datos oficiales de tipo estructurado y más de 21.000 recursos totales, sumando la capa cartográfica federada, las visualizaciones, las historias y los activos de la comunidad. El catálogo se distribuye en 29 categorías temáticas, con una concentración marcada en función pública, educación y salud. La reutilización se agrupa en un grupo reducido de conjuntos asociados a la pandemia, la contratación estatal, la tasa de cambio, los medicamentos y los programas sociales. Ese contraste entre un catálogo amplio y una reutilización concentrada orienta las prioridades de la estrategia.

La estrategia se organiza en tres ejes que estructuran y atraviesan las diez secciones del documento. El primero eleva la calidad de los datos con potencial para inteligencia artificial hasta un nivel verificable mediante una autoevaluación sencilla. El segundo enriquece los metadatos y crea una sección especializada en el portal para que un desarrollador encuentre, entienda y reutilice esos datos adecuadamente. El tercero busca formar y sostener una comunidad de desarrolladores mediante un programa de capacitación y un modelo de gobernanza con indicadores.

La estrategia responde a la acción 2.20 del CONPES 4144 de 2025, relativa a la formulación de la Estrategia de Datos Abiertos de Colombia, y se articula de manera complementaria con la acción 2.16, que aborda la disponibilidad y calidad de los datos para el desarrollo de sistemas de inteligencia artificial. El documento entrega una autoevaluación de aptitud para IA, lineamientos de metadatos y catálogos, el diseño funcional de una sección de IA en datos.gov.co, un programa de formación para desarrolladores y un esquema de gobernanza, sostenibilidad e indicadores.

Línea CONPES 4144 de 2025: Eje 2, Datos e Infraestructura, acción 2.20.

2. Introducción y visión estratégica

2.1. Propósito de la estrategia

El presente documento formula la Estrategia de Datos Abiertos de Colombia orientada al desarrollo y la adopción de sistemas de inteligencia artificial, en cumplimiento de la acción 2.20 del Documento CONPES 4144 de 2025 y en desarrollo del convenio entre el Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC) y la Universidad de Cartagena. Su propósito consiste en que cualquier entidad pública encuentre orientaciones prácticas para publicar y mejorar sus datos, y que cualquier reutilizador, sea un investigador, un desarrollador, una empresa o un ciudadano, sepa cómo localizarlos, evaluarlos y convertirlos en soluciones basadas en inteligencia artificial.

El punto de partida es un principio operativo verificable, donde la calidad, la disponibilidad y la aptitud técnica de los datos determinan el límite superior del desempeño de cualquier modelo que el Estado construya, contrate o supervise. Un conjunto bien descrito, actualizado y con licencia clara habilita analítica avanzada y amplía el margen de lo que un modelo puede lograr; un conjunto desactualizado o sin diccionario de datos condena al proyecto a una etapa de depuración costosa o, en el peor de los casos, a resultados que no deben usarse para decidir. De ese principio se desprende la lógica de todo el documento, que acompaña el ciclo completo del dato desde su identificación en la entidad productora hasta su reutilización responsable en un sistema de inteligencia artificial.

La estrategia toma como referente metodológico la guía para el uso y aprovechamiento de Datos Abiertos en Colombia, que orienta paso a paso a los actores del ecosistema, y adopta ese mismo enfoque para el caso específico de la inteligencia artificial. Los cuatro productos definidos en el hito del CONPES, que corresponden al índice de calidad, al fortalecimiento de metadatos y catálogos, a la sección de inteligencia artificial del portal y al programa de capacitación para desarrolladores, se integran como instrumentos dentro de los capítulos de la guía y conservan su trazabilidad completa, como muestra la Tabla 1.

2.2. Alcance y público objetivo

El documento se dirige a dos audiencias que participan del mismo ecosistema. La primera la integran las entidades públicas del orden nacional y territorial obligadas por la Ley 1712 de 2014, cuyos equipos directivos, líderes de tecnología y responsables de publicación encontrarán en el capítulo 9 una ruta completa para preparar sus datos. La segunda audiencia reúne a los reutilizadores, es decir, a la academia, los investigadores,

los desarrolladores, las empresas, los emprendedores y la sociedad civil, que disponen en el capítulo 10 de una guía metodológica para transformar datos abiertos en sistemas de inteligencia artificial responsables. Los capítulos restantes ofrecen el contexto, el marco normativo, el modelo estratégico y los instrumentos técnicos que sostienen ambos recorridos.

La Tabla 1 presenta la correspondencia entre los cuatro componentes y los capítulos que los desarrollan, de modo que el cumplimiento del hito del CONPES resulte verificable sin que los productos constituyan el eje de la exposición.

Tabla 1. Correspondencia entre los componentes del Entregable 8 y los capítulos de la estrategia

Componente del Entregable 8 (anexo técnico)	Producto o evidencia	Estado de cumplimiento	Capítulos que lo desarrollan
i) Índice de calidad de datos abiertos publicados en datos.gov.co, basado en la autoevaluación de aptitud para IA	Índice de Aptitud para Inteligencia Artificial (IAA), integrado por la lista de verificación de ocho aspectos (Tabla 16), la fórmula de cálculo y la escala de cuatro niveles (Tabla 17). Aplicado sobre diecisiete conjuntos priorizados del catálogo (Tabla 10) y sobre los ocho criterios mínimos de la sección 6.1	Cumplido. Instrumento definido, con fórmula y escala, y aplicado sobre una muestra real del catálogo	Capítulo 9 (paso 3) y capítulo 6 (criterios mínimos)
ii) Fortalecimiento de metadatos y catálogos de datos para el uso de datos abiertos en sistemas de IA	Diagnóstico del esquema de metadatos vigente del portal (sección 7.1) y estándar de quince campos mínimos con su estado actual y la acción propuesta para cada uno (Tabla 12), acompañado de los lineamientos de vocabularios controlados e identificadores persistentes (sección 7.3) y de la ruta de automatización del levantamiento (sección 7.4)	Cumplido. La incorporación de los cinco campos nuevos al esquema del portal corresponde al MinTIC en la fase de implementación	Capítulo 7
iii) Sección del portal datos.gov.co orientada a soluciones y casos de uso de IA basados en datos abiertos	Diseño funcional de la sección, con ocho componentes y el insumo de metadatos que consume cada uno (Tabla 13), criterio de ingreso de conjuntos, ficha normalizada de caso de uso (Anexo D) y cinco casos nacionales con componente de inteligencia artificial ya publicados (Tabla 21)	Cumplido en el alcance del entregable, que es el diseño funcional. La implementación exige los permisos y definiciones técnicas relacionados al cierre de la sección 7.5	Capítulo 7 (sección 7.5) y capítulo 10
iv) Programa de capacitación en datos abiertos para desarrolladores de IA	Programa de siete temas que reproducen el ciclo de vida del dato, con contenidos y justificación de cada uno (Tabla 15), ruta complementaria de formación para entidades productoras (Tabla 14) y esquema	Cumplido. La apertura de cohortes y la certificación de participantes son posteriores a la entrega	Capítulo 8

Componente del Entregable 8 (anexo técnico)	Producto o evidencia	Estado de cumplimiento	Capítulos que lo desarrollan
	de certificación por competencias con comunidad de práctica (sección 8.3)		

Fuente: elaboración propia a partir del anexo técnico del proyecto IA para el Estado.

La correspondencia general de la Tabla 1 se desarrolla en el Anexo B, cuya matriz de trazabilidad descompone los cuatro componentes del entregable en dieciocho obligaciones verificables y las vincula con las acciones 2.16 y 2.20 del CONPES 4144 de 2025, con las exigencias del marco normativo aplicable y con los indicadores de seguimiento.

Es importante mencionar que, el índice de calidad se materializa en este documento como el Índice de Aptitud para Inteligencia Artificial (IAA), que los capítulos 6 y 9 desarrollan en sus criterios, su fórmula y su escala de niveles.

3. Marco normativo y de política pública

La estrategia se inscribe en el ecosistema de política pública de datos y de gobierno digital del país. El Documento CONPES 4144 de 2025 establece la Política Nacional de Inteligencia Artificial con una hoja de ruta hacia 2030, y su acción 2.20 vincula de manera directa con la estrategia de datos abiertos. El presente documento operativiza esa acción para el componente de datos abiertos y aporta instrumentos concretos de medición y publicación.

La Ley 1712 de 2014, Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional, constituye el cimiento del ecosistema. La ley consagra el principio de divulgación proactiva, define a los sujetos obligados y ordena publicar la información en formatos accesibles y reutilizables, con lo cual convierte la apertura en un deber y no en una concesión discrecional. De su artículo sexto se deriva la definición legal de datos abiertos que rige en el país, entendidos como datos primarios sin procesar, en formatos estándar e interoperables, puestos a disposición de cualquier ciudadano dispuesta para ser reutilizados y apoyar la creación de servicios derivados (Congreso de la República de Colombia, 2014).

La Resolución 1519 de 2020 del MinTIC fija las condiciones de publicación de información, accesibilidad web, seguridad digital y datos abiertos para los sujetos obligados por la Ley 1712 de 2014, y su anexo de datos abiertos define los estándares mínimos de estructura, formato y actualización que la estrategia adopta como base técnica (MinTIC, 2020). Sobre ese cimiento normativo se articulan el Plan Nacional de Infraestructura de Datos (PNID), la Infraestructura de Datos del Estado Colombiano (IDEC), el Modelo de Gobierno Digital y la Guía de estándares de calidad e interoperabilidad de datos abiertos del ministerio, que aportan el lenguaje común de interoperabilidad y calidad sobre el que se construyen los lineamientos de las secciones 6 y 7.

La Política de Gobierno Digital, establecida por el Decreto 767 de 2022, sitúa la apertura de datos dentro de una agenda mayor de transformación digital del Estado. El decreto define la política como el uso estratégico de las tecnologías de la información para mejorar la calidad de vida de los ciudadanos y la competitividad del país, y organiza sus habilitadores, entre ellos la arquitectura, la seguridad y los servicios ciudadanos digitales, sobre los que descansa la gestión del dato como activo institucional. La publicación de datos abiertos aptos para inteligencia artificial materializa dos de sus propósitos, la toma de decisiones basada en datos y el impulso a la innovación pública (MinTIC, 2022).

Además, el documento CONPES 4144 de 2025, y su acción 2.20, ordenan publicar una estrategia de datos abiertos que fortalezca la apertura y el aprovechamiento de datos

para el desarrollo de sistemas de inteligencia artificial, y su acción 2.16 vincula la disponibilidad de datos abiertos de calidad con la viabilidad de esos sistemas (Departamento Nacional de Planeación, 2025). El presente documento operativiza ambas acciones para el componente de datos abiertos y aporta los instrumentos de medición, publicación y formación que las entidades y los reutilizadores necesitan.

El Plan Nacional de Infraestructura de Datos y la Infraestructura de Datos del Estado Colombiano completan el cuadro. El plan, expedido mediante la Resolución 460 de 2022 y construido de manera conjunta por el MinTIC, el Departamento Nacional de Planeación y la Presidencia de la República, establece que la disponibilidad, la calidad, la interoperabilidad y la reutilización de los datos son condiciones habilitantes para el desarrollo de soluciones de inteligencia artificial confiables, y la infraestructura aporta el lenguaje común de gobernanza y de intercambio sobre el que se construyen los lineamientos de metadatos del capítulo 7 (MinTIC et al., 2022). En el plano ético, la Guía Ética para la Implementación, Desarrollo y Uso de Sistemas de Inteligencia Artificial en Entidades Públicas de Colombia define los principios que el capítulo 5 adopta como marco de actuación (Gobierno de Colombia, 2025). La Tabla 2 sintetiza el aporte de cada instrumento a la estrategia.

Tabla 2. Instrumentos normativos y de política, y su aporte a la estrategia

Instrumento	Contenido pertinente	Aporte a esta guía
Ley 1712 de 2014	Divulgación proactiva, sujetos obligados y definición legal de datos abiertos	Fundamento del deber de publicar en formatos reutilizables (capítulo 9)
Resolución 1519 de 2020	Estándares de publicación, accesibilidad, seguridad y anexo de datos abiertos	Base técnica del paso de publicación y del seguimiento anual (capítulo 9)
Decreto 767 de 2022 (Política de Gobierno Digital)	Gestión del dato como activo y decisiones basadas en datos	Marco institucional de los pasos de gobernanza y seguimiento (capítulo 9)
CONPES 4144 de 2025	Política Nacional de IA; acciones 2.16 y 2.20	Mandato directo de la estrategia y de sus cuatro instrumentos
PNID e Infraestructura de Datos del Estado	Condiciones habilitantes de calidad e interoperabilidad	Lenguaje común de los estándares de metadatos y catálogos (capítulo 7)
Guía Ética de IA para entidades públicas	Principios y lineamientos éticos del ciclo de vida de la IA	Principios del modelo estratégico y del desarrollo responsable (capítulos 5 y 6)

Fuente: elaboración propia.

La articulación con estos instrumentos aprovecha el trabajo existente y garantiza que los productos de la estrategia, es decir, la autoevaluación de aptitud para IA, el estándar de metadatos y la sección de IA del portal, resulten compatibles con las obligaciones vigentes de las entidades y con las metas de la política de inteligencia artificial. La estrategia también toma como referencia principios internacionales de datos abiertos y

los principios FAIR de datos localizables, accesibles, interoperables y reutilizables, que ofrecen un marco técnico maduro para el caso de uso de la inteligencia artificial (Wilkinson et al., 2016). La correspondencia entre cada exigencia normativa y el instrumento de la estrategia que la atiende se presenta en la Tabla B3 del Anexo B.

4. Contexto

4.1. Tendencias internacionales

Colombia llega a esta estrategia con una posición internacional favorable y con una debilidad localizada que el documento se propone corregir. En la edición 2025 del índice OURdata de la Organización para la Cooperación y el Desarrollo Económicos, que mide las políticas de datos gubernamentales abiertos, el país obtuvo una calificación general de 0,68 sobre 1, por encima del promedio de los países miembros, que fue de 0,53 (OECD, 2026). El resultado por pilares revela dónde está el trabajo pendiente. El país sobresale en disponibilidad de datos y en apoyo gubernamental a la reutilización, y queda por debajo del promedio en accesibilidad de los datos, dimensión asociada a los formatos abiertos y a los estándares de metadatos. La Tabla 3 resume estos resultados.

Tabla 3. Resultados de Colombia en el índice OURdata 2025 de la OCDE

Pilar	Colombia	Promedio OCDE	Lectura para la estrategia
Disponibilidad de datos	0,71	0,53	La oferta de datos es amplia; el reto no es publicar más, sino publicar mejor
Accesibilidad de los datos	0,60	0,67	Única dimensión bajo el promedio; exige mejorar formatos, metadatos y catálogos, que concentran los capítulos 6 y 7
Apoyo a la reutilización	0,75	0,40	La institucionalidad de fomento existe; la guía para reutilizadores del capítulo 10 la capitaliza
Índice general	0,68	0,53	Posición favorable con una brecha localizada y corregible

Fuente: elaboración propia con datos de OCDE (2026).

La brecha del país no está en la cantidad de datos publicados ni en la voluntad institucional de promover su uso, sino en la capa técnica que hace que un dato sea procesable por una máquina sin fricción, es decir, en los formatos, los metadatos y la descubribilidad. Esa capa es justo la que condiciona la viabilidad de los sistemas de inteligencia artificial, y por eso los instrumentos centrales de esta estrategia se concentran en ella. La segunda edición del Global Data Barometer, que evalúa la gobernanza, las capacidades y el uso de los datos para el bien público, respalda esa lectura para la región, pues encuentra en América Latina una base institucional sólida junto a tareas pendientes en calidad, disponibilidad y uso de los datos, y advierte que pocos países cuentan con políticas claras para orientar el uso ético de la inteligencia artificial (Global Data Barometer, 2025). Colombia dispone de esa política desde la expedición de la Guía Ética y del CONPES 4144, y la presente estrategia completa el

eslabón que la medición internacional señala como faltante, el de los datos de calidad para sistemas de IA.

Los marcos técnicos internacionales ofrecen el lenguaje común para cerrar esa brecha. Los principios FAIR piden datos localizables, accesibles, interoperables y reutilizables (Wilkinson et al., 2016), la Carta Internacional de Datos Abiertos establece la apertura por defecto, la oportunidad y la interoperabilidad como atributos de la publicación (Open Data Charter, 2015), y el vocabulario DCAT del Consorcio de la Web normaliza la descripción de catálogos (World Wide Web Consortium, 2014). La estrategia adopta estos referentes en sus instrumentos y los traduce a pasos operativos en los capítulos 7 y 9.

4.2. Diagnóstico del catálogo nacional de datos abiertos para IA

El portal aloja 8.422 conjuntos de datos oficiales de tipo estructurado, que constituyen el núcleo tabular con acceso programático y, por tanto, la base analítica sobre la que puede construirse inteligencia artificial. A ese núcleo se suman una extensa capa cartográfica federada, proveniente en buena medida del Instituto Geográfico Agustín Codazzi y de portales territoriales, 838 visualizaciones entre gráficos y mapas, 322 historias y reportes, y 1.421 activos creados por la comunidad. La actividad de los usuarios no institucionales se concentra en vistas y visualizaciones derivadas, mientras que los conjuntos de datos propiamente dichos son en su totalidad de origen oficial. La Tabla 4 resume la composición del catálogo.

Tabla 4. Composición del catálogo de datos.gov.co por tipo de activo

Tipo de activo	Método de obtención	Cantidad	Observación
Conjuntos de datos oficiales	Consulta al catálogo de descubrimiento de Socrata sobre el dominio www.datos.gov.co con el filtro de tipo de activo dataset. El conteo se toma del campo resultSetSize de la respuesta, con el parámetro limit en cero para recuperar el total sin descargar registros.	8.422	Núcleo tabular con acceso por API y base analítica para IA
Capa cartográfica federada	Conteo de los recursos que el portal expone por federación desde dominios externos, obtenido en el navegador del catálogo con el filtro de recursos externos. La cifra se reporta como cota inferior, porque el catálogo de descubrimiento trunca el conteo total en 10.000 registros y porque la disponibilidad de los geoservicios federados varía entre consultas.	> 10.000	Predominan mosaicos cartográficos, con menor aporte a la IA tabular

Tipo de activo	Método de obtención	Cantidad	Observación
Visualizaciones (gráficos y mapas)	Suma de dos consultas independientes al catálogo de descubrimiento sobre el mismo dominio, con los filtros de tipo de activo chart y map, que arrojan 703 gráficos y 135 mapas.	838	703 gráficos y 135 mapas derivados de los conjuntos
Historias y reportes	Consulta al catálogo de descubrimiento con el filtro de tipo de activo story, que agrupa las narrativas y los tableros publicados por el ministerio, entre ellos los Reportes del Portal.	322	Narrativas y tableros, incluidos los Reportes del Portal
Activos de la comunidad	Consulta al catálogo de descubrimiento con el filtro de procedencia community, que aísla los activos creados por usuarios no institucionales sobre conjuntos oficiales, en su mayoría vistas filtradas y visualizaciones derivadas.	1.421	Vistas creadas por usuarios, con calidad por validar

Fuente: elaboración propia con datos del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

La composición aporta un matiz que orienta toda la estrategia, donde una parte considerable de los recursos del portal corresponde a mosaicos cartográficos y a productos de cumplimiento de la Ley 1712 de 2014, como el registro de activos de información, las tablas de retención documental, los esquemas de publicación y los índices de información clasificada y reservada, que figuran entre las etiquetas más frecuentes del catálogo (Figura 1). Estos activos cumplen una función de transparencia valiosa por sí misma. El conjunto realmente apto para inteligencia artificial forma un subconjunto del núcleo oficial, lo que confirma el valor de priorizar y concentrar los esfuerzos donde el retorno es mayor.

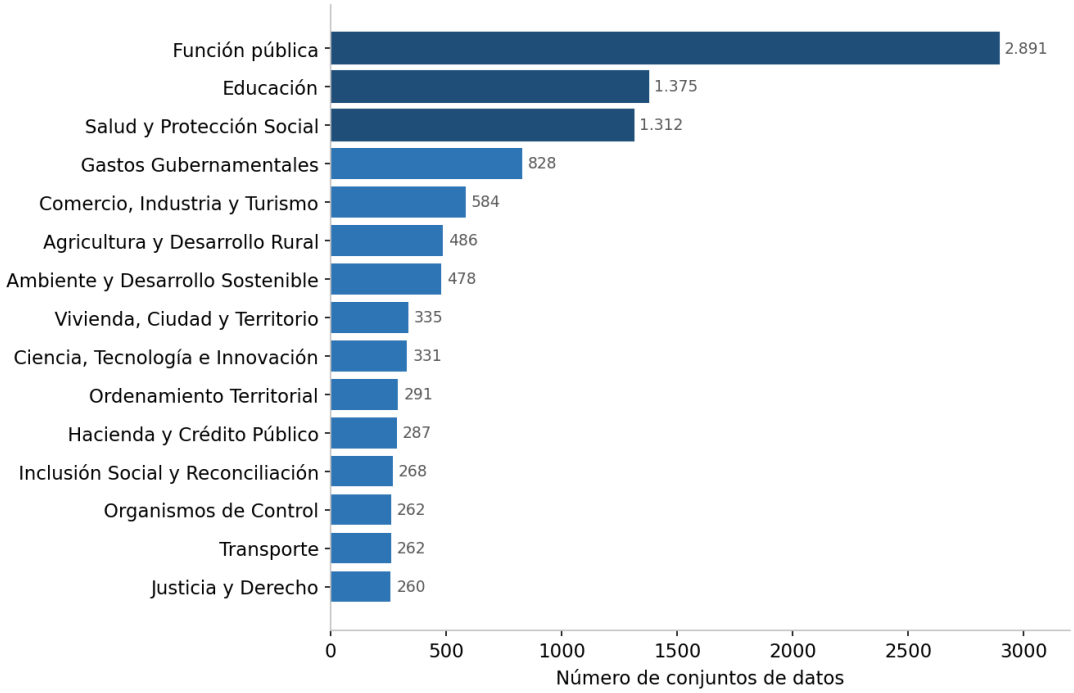


Figura 1. Distribución temática del catálogo de datos.gov.co (15 categorías principales)

Fuente: elaboración propia con datos del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

La categoría de función pública encabeza el catálogo con 2.891 conjuntos, seguida por educación con 1.375 y salud y protección social con 1.312. Estas tres categorías reúnen cerca de dos terceras partes de todos los conjuntos estructurados. La concentración temática admite una lectura estratégica directa. Los dominios con mayor oferta de datos coinciden en buena medida con aquellos donde la inteligencia artificial pública ofrece casos de uso de alto impacto, como la analítica de contratación, la vigilancia epidemiológica o el análisis de trayectorias educativas.

4.2.1 Concentración y madurez de la reutilización

El uso del portal concentra su visualización en un grupo reducido de datos. La Figura 2 ordena los diez conjuntos más consultados por número acumulado de visualizaciones y muestra cinco frentes de demanda, que abarcan la información epidemiológica de la pandemia, la contratación estatal a través del SECOP, la tasa de cambio representativa del mercado, el código único de medicamentos y los programas sociales.

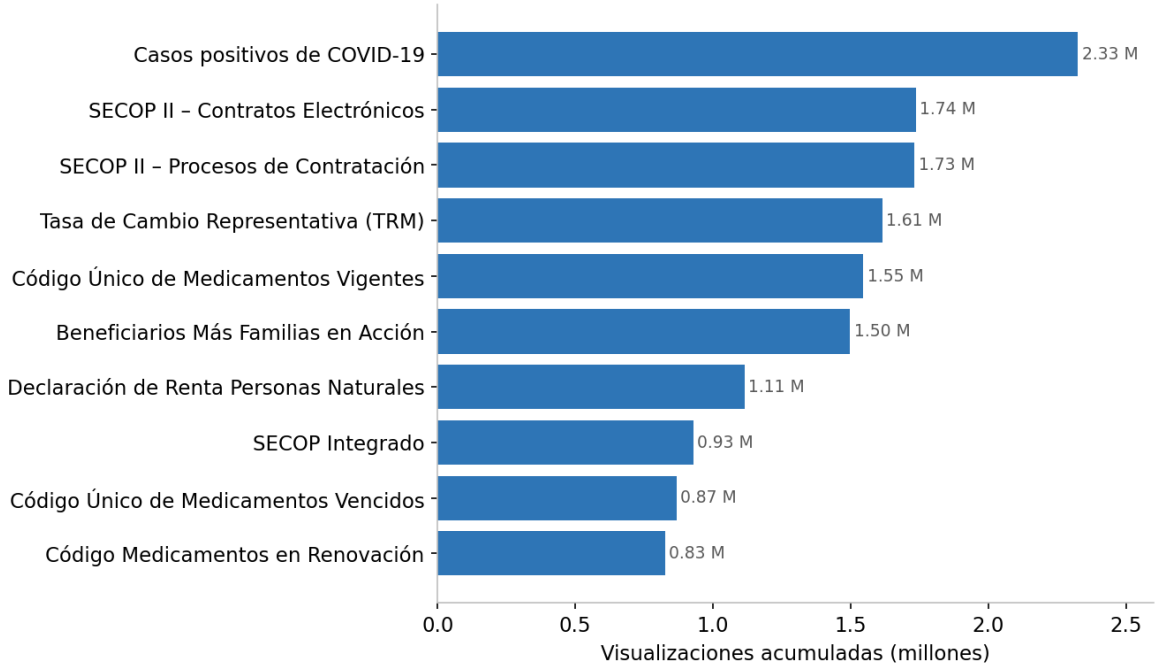


Figura 2. Conjuntos de datos más consultados en datos.gov.co

Fuente: elaboración propia con métricas de visualización del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

El conjunto de casos positivos de COVID-19 supera los 2,3 millones de visualizaciones y acumula más de un millón de descargas, seguido por los dos conjuntos del SECOP, que juntos rozan los 4,1 millones de vistas. La combinación de visualizaciones y descargas confirma que la demanda de datos para analítica ya existe y se orienta hacia información de alto valor cívico y económico. De ahí surge un criterio de priorización orientado a dirigir los esfuerzos de calidad hacia los dominios que la ciudadanía y que los desarrolladores ya consultan, y a su vez, activar sectores con datos valiosos y baja utilización.

Cabe destacar que la metadata del portal distingue tres marcas de tiempo por conjunto, que son la fecha de creación, la fecha de última actualización de los datos y la fecha de última actualización de los metadatos. Esa distinción permite observar la vigencia real de la información.

4.2.2 Formatos, acceso y licenciamiento

El acceso constituye la principal fortaleza del portal. Cada conjunto oficial alojado en la plataforma Socrata se expone de forma automática mediante una interfaz de programación y en formatos estructurados como CSV, JSON, XLSX, XML y GeoJSON, lo que lo hace legible por máquina de manera directa. Esta característica sitúa al núcleo oficial en una posición favorable para la inteligencia artificial. La capa federada ofrece un contraste, con recursos en PDF y hojas de cálculo de estructura irregular que requieren una etapa adicional de extracción.

La licencia es un campo nativo del portal cuya declaración no es obligatoria en la configuración actual. La estrategia propone elevarla a requisito para los conjuntos que aspiren a la priorización, de manera que la exigencia opera como criterio de elegibilidad dentro del esquema de priorización y no como una obligación general para la totalidad del catálogo.

4.3. Retos actuales

El catálogo combina fortalezas de cobertura y acceso con oportunidades de mejora en aspectos técnicos y en intervenciones puntuales que elevan la aptitud de los conjuntos con potencial analítico.

La capa federada ofrece un contraste, con recursos en PDF y hojas de cálculo de estructura irregular que requieren una etapa adicional de extracción. La vigencia merece atención particular, porque la metadata del portal distingue la fecha de actualización de los datos de la fecha de actualización de los metadatos, y algunos conjuntos muestran metadatos recientes junto a datos obtenidos años atrás pese a declarar frecuencias diarias o trimestrales. La licencia es un campo nativo del portal, aunque su declaración avanza de forma desigual en el catálogo.

Dos de los retos observados tienen respaldo en la literatura especializada: el diccionario de datos incompleto y la brecha entre política declarada e implementación. El diagnóstico encontró, en conjuntos de consulta masiva, fichas que no documentan la motivación, la composición, el proceso de recolección y los usos recomendados de cada conjunto antes de su distribución, elementos valorados para el desarrollo de modelos de aprendizaje automático (Gebru et al., 2021); el estándar de metadatos adopta esa lógica. La distancia entre la política declarada y su implementación efectiva, visible en frecuencias de actualización que no se cumplen, coincide con lo que la investigación comparada de políticas de datos abiertos identifica como el eslabón débil de estas iniciativas, la ausencia de seguimiento con indicadores sobre el uso real de los datos (Zuiderwijk y Janssen, 2014).

La Tabla 5 resume cada aspecto, la evidencia observada en el portal y su implicación práctica, y el anexo A documenta cada uno con conjuntos de datos vigentes, entidad responsable y enlace directo.

Tabla 5. Aspectos por fortalecer, evidencia observada en el portal e implicación para la IA

Aspecto por fortalecer	Evidencia observada en el portal	Implicación para la IA
Vigencia de los datos	El conjunto de COVID-19 mantiene datos de enero de 2024 con metadatos recientes y frecuencia declarada diaria	Conviene validar la vigencia efectiva antes de entrenar
Legibilidad por máquina	Recursos de la capa federada en PDF y hojas de cálculo de estructura irregular	Requiere una etapa de extracción y estructuración
Consistencia de metadatos	Descripciones de columnas por completar y campos de entidad que difieren de la atribución del conjunto	Se recomienda depurar la descripción y la atribución
Uniformidad del licenciamiento	La licencia es un campo nativo y una parte del catálogo la declara de forma explícita	Conviene declarar la licencia abierta en todos los conjuntos
Foco analítico del catálogo	Amplia presencia de mosaicos cartográficos y productos de cumplimiento (RAI, tablas de retención, esquemas de publicación)	Orienta la priorización hacia el núcleo analítico
Curaduría de activos comunitarios	1.421 vistas creadas por usuarios no institucionales	Sugiere una validación institucional antes de su uso
Ampliación de la reutilización	Diez conjuntos concentran millones de vistas y descargas	Abre la oportunidad de activar datos poco utilizados

Fuente: elaboración propia a partir del diagnóstico del catálogo y de la metadata técnica de datos.gov.co. La evidencia detallada se presenta en el anexo A.

El diagnóstico se completa con insumos que reposan en las áreas técnicas del ministerio. Se solicita a MinTIC el inventario actualizado de conjuntos de datos con sus responsables, las métricas de uso y descarga de datos.gov.co desagregadas por conjunto, los reportes de calidad disponibles, la relación de datasets priorizados por las entidades y los criterios de selección aplicados.

4.4. Síntesis DOFA articulada a los ejes estratégicos

Los hallazgos del diagnóstico se sintetizan en una matriz DOFA que ordena las fortalezas y debilidades internas del ecosistema del portal junto a las oportunidades y amenazas de su entorno. La matriz cumple una función de articulación, pues cada elemento se vincula con el eje estratégico que lo atiende o lo capitaliza, de manera que el modelo del capítulo 5 no parte de postulados abstractos sino de la evidencia documentada en las secciones anteriores y en el anexo A. La Tabla 6 presenta la síntesis.

Tabla 6. Matriz DOFA del diagnóstico y su articulación con los ejes estratégicos

Cuadrante	Elemento del diagnóstico	Eje que lo atiende
Fortalezas	F1. Acceso programático nativo por API y formatos estructurados para todo el núcleo oficial de 8.422 conjuntos	E2 lo capitaliza

	F2. Cobertura temática amplia en 29 categorías, con concentración en dominios de alto impacto para IA (función pública, educación, salud)	E1 lo capitaliza
	F3. Demanda comprobada, con millones de visualizaciones y descargas en contratación, salud, tasa de cambio y programas sociales	E3 lo capitaliza
	F4. Desempeño sobre el promedio OCDE en disponibilidad de datos (0,71) y en apoyo a la reutilización (0,75)	E3 lo capitaliza
Debilidades	D1. Accesibilidad bajo el promedio OCDE (0,60 frente a 0,67), asociada a formatos y estándares de metadatos	E2 la corrige
	D2. Diccionarios de datos incompletos y metadatos inconsistentes, incluso en conjuntos de consulta masiva	E2 la corrige
	D3. Vigencia comprometida, con series detenidas bajo metadatos recientes y frecuencias declaradas sin cumplir	E1 la corrige
	D4. Licenciamiento desigual, con conjuntos sin licencia abierta declarada	E1 la corrige
	D5. Reutilización concentrada en un grupo reducido de conjuntos, con sectores valiosos subutilizados	E3 la corrige
Oportunidades	O1. Mandato vigente del CONPES 4144 y de la Guía Ética, que dan respaldo de política a la estrategia	E1, E2 y E3
	O2. Oferta formativa activa de la universidad corporativa del MinTIC y ecosistema de casos de uso con IA en crecimiento	E3 la aprovecha
	O3. Referentes internacionales maduros, como el reglamento europeo de datos de alto valor y las mediciones del Global Data Barometer	E1 y E2
	O4. Automatización disponible (rastreador del proyecto MinTIC y Universidad de Cartagena) para el levantamiento de metadatos a escala	E2 la aprovecha
Amenazas	A1. Desactualización sostenida de conjuntos emblemáticos, que erosiona la confianza de los reutilizadores	E1 la mitiga
	A2. Dependencia de plataformas federadas externas cuya continuidad no controla el portal	E2 la mitiga
	A3. Riesgos de reidentificación y de sesgos en dominios sensibles como salud, educación y protección social	E1 la mitiga
	A4. Rotación de personal en las entidades y pérdida de capacidades al término del convenio	E3 la mitiga

Fuente: elaboración propia a partir del diagnóstico de las secciones 4.2 y 4.3, del anexo A y de OCDE (2026).

La lectura cruzada de la matriz explica la arquitectura de la estrategia. Las combinaciones ofensivas, que usan fortalezas para aprovechar oportunidades, sostienen los ejes E2 y E3, porque el acceso programático nativo y la demanda comprobada permiten capitalizar la oferta formativa y el ecosistema de casos para ampliar la reutilización.

Las combinaciones adaptativas, que corrigen debilidades con ayuda de las oportunidades, concentran al eje E2, dado que la brecha de accesibilidad y los diccionarios incompletos se cierran con el estándar de metadatos apoyado en la automatización disponible y en los referentes internacionales.

Las combinaciones defensivas, que enfrentan amenazas desde las fortalezas, corresponden al eje E1, pues la autoevaluación de aptitud y la verificación de vigencia protegen la confianza que la demanda ya construyó. Las combinaciones de supervivencia, que atienden a la vez debilidades y amenazas, recaen en el eje E3, cuyo modelo de gobernanza con roles, indicadores y formación de formadores responde a la concentración de la reutilización y a la pérdida de capacidades tras el convenio. Cada acción de los capítulos 5 a 9 puede rastrearse hasta un cuadrante de la Tabla 6 (matriz DOFA), lo que da al lector institucional un criterio para evaluar la pertinencia de la estrategia frente a su propio contexto.

5. Modelo estratégico de datos abiertos para inteligencia artificial

5.1. Principios

La estrategia adopta como marco de actuación los principios de la Guía Ética para la Implementación, Desarrollo y Uso de Sistemas de Inteligencia Artificial en Entidades Públicas de Colombia, construida con referentes de la OCDE, la Unesco, las Naciones Unidas y el Alan Turing Institute (Gobierno de Colombia, 2025). La guía establece la transparencia, la explicación, la privacidad, el control humano de las decisiones, la seguridad, la responsabilidad, la no discriminación, la inclusión, la prevalencia de los derechos de niñas, niños y adolescentes y el beneficio social como principios del ciclo de vida de los sistemas de inteligencia artificial en el sector público. Para el dominio de los datos abiertos, estos principios se traducen en obligaciones concretas. La transparencia exige documentar el origen y el método de recolección de cada conjunto; la privacidad obliga a verificar la ausencia de datos personales antes de publicar; la no discriminación pide declarar los sesgos conocidos de los datos; y el control humano supone que los resultados de un modelo entrenado con datos abiertos apoyan la decisión pública sin reemplazarla.

A esos principios éticos se suman los principios técnicos FAIR, que piden datos localizables, accesibles, interoperables y reutilizables (Wilkinson et al., 2016), y los de la Carta Internacional de Datos Abiertos, en particular la apertura por defecto y la oportunidad de la publicación (Open Data Charter, 2015). La combinación de ambos planos, el ético y el técnico, define el estándar de esta estrategia. Para efectos de esta estrategia, un dato abierto es apto para inteligencia artificial cuando cumple simultáneamente condiciones técnicas de procesamiento automatizado y condiciones éticas y jurídicas de publicación verificables.

5.2. Objetivos y ejes

El objetivo general de la estrategia consiste en fortalecer la apertura y el aprovechamiento de los datos abiertos de Colombia para el desarrollo de sistemas de inteligencia artificial, en cumplimiento de la acción 2.20 del CONPES 4144 de 2025. Ese objetivo se despliega en tres ejes que organizan los instrumentos del documento. El primero eleva la calidad de los datos con potencial para inteligencia artificial hasta un nivel verificable mediante una autoevaluación sencilla. El segundo enriquece los metadatos y crea una sección especializada en el portal para que un desarrollador encuentre, entienda y reutilice esos datos. El tercero busca formar y sostener una comunidad de práctica mediante rutas de capacitación y un modelo de gobernanza con

indicadores. La Tabla 7 presenta los ejes, los capítulos que los despliegan y el resultado esperado de cada uno.

Tabla 7. Ejes estratégicos, capítulos que los despliegan y resultados esperados

Eje	Enfoque	Capítulos	Resultado esperado (indicador)
E1. Calidad y priorización	Elevar a un umbral medible la calidad de los datos con potencial para IA	6 y 9	Conjuntos en nivel Listo para IA (autoevaluación)
E2. Metadatos y descubribilidad	Enriquecer metadatos y crear la sección de IA del portal	7 y 10	Catálogo semántico y sección de IA operativa
E3. Capacitación y gobernanza	Formar desarrolladores y sostener el modelo con indicadores	8	Comunidad certificada y tablero de indicadores activo

Fuente: elaboración propia.

Los tres ejes operan de forma encadenada. El eje E1 busca asegurar la existencia de datos aptos, el eje E2 procura que sean encontrables y comprensibles, y el eje E3 promueve su uso y la continuidad del proceso. Cada eje potencia al siguiente, y los capítulos de guía traducen sus instrumentos en pasos que las entidades y los reutilizadores pueden ejecutar con los recursos disponibles. La cadena completa de trazabilidad entre ejes, instrumentos e indicadores se documenta en la Tabla B4 del Anexo B, que remite a los códigos de indicador desarrollados en el Anexo E.

5.3. Ecosistema de actores

Un ecosistema sostenible de datos abiertos e inteligencia artificial requiere que cada actor entienda su papel y el valor que recibe a cambio de su participación (Janssen et al., 2012). Las entidades públicas producen y publican los datos, y obtienen a cambio analítica para sus propias decisiones, cumplimiento normativo verificable y visibilidad de su gestión. El MinTIC coordina la política, administra el portal y consolida los indicadores. La academia y los investigadores aportan métodos, forman talento y validan la calidad de los datos con su uso; los casos publicados en el portal muestran que universidades y grupos de investigación figuran entre los reutilizadores más activos. Los desarrolladores, las empresas y los emprendedores convierten los datos en productos y servicios, desde visualizaciones hasta plataformas predictivas, y su demanda orienta la priorización de la apertura. La sociedad civil y el periodismo de datos ejercen control social sobre la gestión pública y amplían la base de usuarios del portal. La Tabla 8 resume estos papeles.

Tabla 8. Actores y papel en el ecosistema

Actor	Papel en el ecosistema
Entidades públicas	Producen, publican, actualizan y mejoran la calidad de sus conjuntos de datos
MinTIC	Coordina la política, administra datos.gov.co y consolida los indicadores
Academia e investigadores	Aportan métodos, forman talento y validan los datos con su uso
Desarrolladores, empresas y emprendedores	Convierten los datos en productos, servicios y modelos
Sociedad civil y periodismo de datos	Ejercen control social y amplían la reutilización cívica

Fuente: elaboración propia.

El ecosistema funciona cuando el ciclo se cierra, es decir, cuando la reutilización orienta a la publicación. Los mecanismos de retroalimentación de esta guía, entre ellos las fichas de caso de uso, la sección de usos del portal y los indicadores de reutilización, existen para indicar a las entidades qué datos generan valor y ajusten su priorización en consecuencia.

6. Datos abiertos para inteligencia artificial

Este capítulo responde la pregunta por la identificación de conjuntos de alto valor para inteligencia artificial y define qué convierte a un dato abierto en un dato listo para IA. Presenta los criterios mínimos, la clasificación que ordena la priorización, una muestra priorizada del catálogo nacional, los riesgos que el uso de estos datos comporta y los aspectos éticos que lo condicionan.

6.1. Datos listos para inteligencia artificial y criterios mínimos

Un conjunto de datos está listo para inteligencia artificial cuando una máquina puede procesarlo sin intervención manual y un equipo puede confiar en sus resultados. Esa condición se verifica con los ocho aspectos de la autoevaluación del capítulo 9, que operan como criterios mínimos, a saber, estructura tabular clara, completitud de las columnas principales, vigencia acorde con la frecuencia declarada, diccionario de datos completo, licencia abierta visible, acceso por enlace estable o interfaz de programación, uso de códigos estándar como la DIVIPOLA y declaración del origen del dato.

Los cuatro niveles de la Tabla 17 gradúan el cumplimiento, y el nivel Listo para IA exige los ocho aspectos a la vez. La exigencia simultánea tiene fundamento técnico, pues un conjunto perfecto en siete aspectos pero desactualizado produce modelos vencidos, y uno vigente pero sin licencia expone al proyecto a un riesgo jurídico que ninguna entidad aceptaría. El cumplimiento simultáneo de esos ocho criterios es lo que mide el Índice de Aptitud para Inteligencia Artificial (IAA) propuesto.

6.2. Alto valor, datos críticos y datos estratégicos

La estrategia distingue tres categorías que orientan la priorización.

- Los datos de alto valor tienen demanda comprobada e impacto económico o social amplio, como los del SECOP, la tasa de cambio o el código único de medicamentos.
- Los datos críticos sostienen decisiones sensibles o servicios esenciales, en dominios como salud, seguridad o inclusión social, donde la precisión del modelo tiene consecuencias directas.
- Los datos estratégicos, con reutilización aún baja, habilitarían casos de uso emergentes de alto retorno, como los datos ambientales para alertas tempranas o los agropecuarios para la gestión del riesgo climático.

La clasificación concentra los esfuerzos de calidad sobre los conjuntos de mayor retorno y se aplica en el paso 1 de la guía para entidades. El enfoque coincide con la práctica

internacional, pues la Unión Europea definió por reglamento seis categorías temáticas de conjuntos de datos de alto valor, entre ellas la geoespacial, la de observación de la Tierra y medio ambiente, la meteorológica, la estadística, la de sociedades y la de movilidad, cuya publicación gratuita, en formato legible por máquina y mediante interfaces de programación es obligatoria para los Estados miembros (Comisión Europea, 2023). La Tabla 9 relaciona los sectores con mayor volumen en el portal y los casos de uso de inteligencia artificial de mayor potencial en cada uno.

Tabla 9. Sectores priorizados y casos de uso de inteligencia artificial de alto potencial

Categoría (volumen en datos.gov.co)	Casos de uso de IA de alto potencial
Función pública (2.891)	Analítica de contratación y detección de anomalías en SECOP, asistentes conversacionales de trámites, priorización de auditorías
Educación (1.375)	Modelos de alerta temprana de deserción, recomendación de oferta educativa, analítica de resultados de pruebas
Salud y protección social (1.312)	Vigilancia epidemiológica, farmacovigilancia sobre el código único de medicamentos, modelos de demanda de servicios
Gastos gubernamentales (828)	Auditoría asistida del gasto, clasificación automática de rubros, análisis de patrones atípicos
Ambiente y desarrollo sostenible (478)	Alertas tempranas, modelos de deforestación y de calidad del aire, análisis de riesgo climático
Agricultura y desarrollo rural (486)	Predicción de cosechas y precios, modelos de riesgo agroclimático, focalización rural
Transporte (262)	Optimización de movilidad, modelos de seguridad vial, análisis de siniestralidad

Fuente: elaboración propia con datos de categorías del catálogo de datos.gov.co.

6.3. Muestra priorizada de conjuntos con aptitud para IA

La aplicación de los criterios mínimos sobre los conjuntos de mayor demanda del catálogo produce la muestra priorizada que presenta la Tabla 10, con diecisiete conjuntos que cubren los sectores priorizados de la Tabla 9, su estado observado y la mejora pendiente para alcanzar la condición de listos para IA. Todos los conjuntos de la muestra declaran licencia abierta Creative Commons BY y BY-SA 4.0, y exponen acceso por interfaz de programación, de modo que las mejoras pendientes se concentran en el diccionario de datos, la vigencia y la atribución de entidad. La muestra documenta el método para que el MinTIC y las entidades lo extiendan al resto del catálogo priorizado, con el apoyo del inventario y de las métricas solicitadas al ministerio en el capítulo 4.

Tabla 10. Muestra priorizada de conjuntos de alto valor y mejora pendiente para la condición de listos para IA

Conjunto (identificador en el portal)	Fortalezas observadas	Mejora pendiente
SECOP II, Procesos de Contratación (p6dx-8zbt)	Demanda masiva, con más de 1,7 millones de vistas y 150 mil descargas; acceso por API; licencia abierta declarada	Completar el diccionario de datos y declarar la cobertura temporal
SECOP II, Contratos Electrónicos (jbjy-vk9h)	Más de 1,7 millones de vistas y 209 mil descargas; acceso por API; licencia abierta declarada	Homologar la atribución de entidad y documentar sesgos de registro
SECOP Integrado (rpmr-utcd)	Más de 940 mil vistas y 66 mil descargas; consolida la contratación de las tres plataformas	Corregir la atribución de entidad, hoy asignada a la Alcaldía de Somondoco en lugar de la ANCP
Casos positivos de COVID-19 en Colombia (gt2j-8ykr)	Conjunto más descargado del portal, con más de un millón de descargas; acceso por API	Declarar de forma explícita el corte real de la serie, detenida desde enero de 2024
Código Único de Medicamentos Vigentes (i7cb-raxc)	Más de 1,5 millones de vistas; referencia nacional del sector salud	Describir 28 de 29 columnas hoy sin diccionario
Beneficiarios Más Familias en Acción (xfif-myr2)	Alta demanda ciudadana, con cerca de 1,5 millones de vistas; cobertura nacional del programa	Restablecer la actualización, detenida frente a la frecuencia trimestral declarada
Estadísticas en educación preescolar, básica y media por departamento (ji8i-4anb)	Serie anual vigente con corte a 2025; diccionario completo; más de 200 mil vistas	Declarar la cobertura temporal y documentar el método de recolección
Resultados únicos Saber Pro (u37r-hjmu)	Referencia nacional de resultados de la educación superior, con demanda sostenida	Restablecer la actualización, detenida desde 2023 frente a la frecuencia semestral declarada, y corregir la atribución de entidad
OVCF, CUIPO, Ejecución de Gastos (4f7r-epif)	Serie trimestral vigente de la Contraloría; diccionario completo en sus 31 columnas; códigos BPIN y MGA	Ampliar la difusión para elevar una reutilización aún baja frente a su valor analítico
Tasa de Cambio Representativa del Mercado (32sa-8pi3)	Actualización diaria efectiva; más de 528 mil descargas; diccionario completo; serie histórica larga	Cumple los criterios de la autoevaluación; candidato inmediato a la condición de listo para IA
Catálogo Nacional de Estaciones del IDEAM (hp9r-jxuu)	Actualización diaria efectiva; diccionario casi completo; base de la red hidrometeorológica nacional	Describir las tres columnas de ubicación que carecen de diccionario
Precipitación (s54a-sgyg)	Registros cada diez minutos con actualización diaria efectiva; diccionario completo; insumo directo de alertas tempranas	Trasladar a la ficha las advertencias de validación que hoy solo aparecen en la descripción
Temperatura Ambiente del Aire (sbwg-7ju4)	Actualización diaria efectiva; serie horaria de cobertura nacional	Diligenciar el diccionario de datos, hoy vacío en sus doce columnas
Evaluaciones Agropecuarias Municipales EVA (2pnw-mmge)	Más de 135 mil vistas y 35 mil descargas; base municipal de producción agrícola con códigos DANE	Actualizar o declarar el corte de la serie 2007 a 2018, detenida desde 2019, y corregir la atribución de entidad
Solicitudes de restitución de tierras por municipio (s87b-tjcc)	Serie mensual vigente; diccionario completo; usa el código DIVIPOLA	Declarar la cobertura temporal en la ficha

Conjunto (identificador en el portal)	Fortalezas observadas	Mejora pendiente
Actores RUNT por municipio y departamento (epfm-5fhb)	Serie mensual vigente; diccionario completo; cobertura nacional	Normalizar los títulos de columnas, que contienen espacios irregulares
Tráfico Portuario Marítimo en Colombia (5r3g-zv5z)	Serie trimestral vigente desde 2018; diccionario completo	Documentar en la ficha el proceso de revisión de cifras que la entidad declara en la descripción

Fuente: elaboración propia con la metadata técnica de datos.gov.co (consultas del 1 y del 16 de julio de 2026); el detalle de evidencias se documenta en el anexo A.

6.4. Riesgos

El uso de datos abiertos en sistemas de inteligencia artificial enfrenta riesgos identificables y gestionables. La formulación de estos riesgos sigue la Guía para la administración del riesgo y el diseño de controles en entidades públicas de Función Pública, en su versión 6, que estructura cada riesgo a partir de su causa y su consecuencia, lo valora con la escala de probabilidad e impacto y exige que cada control declare su propósito y su responsable (Departamento Administrativo de la Función Pública [DAFP], 2022). Los seis riesgos de la Tabla 11 corresponden a las tipologías de gestión y de seguridad digital que la guía distingue, y su valoración se sustenta en la evidencia del diagnóstico del capítulo 4 y del Anexo A. Los controles remiten a instrumentos que la propia estrategia ya define, sin exigir procedimientos adicionales a las entidades ni a los reutilizadores

Tabla 11. Riesgos del uso de datos abiertos en sistemas de inteligencia artificial, valoración y controles

Riesgo	Causa	Consecuencia sobre el sistema de IA	Valoración inherente	Control, propósito y responsable
R1. Vigencia de los datos	La entidad actualiza los metadatos del conjunto sin que la serie avance, con lo que el portal muestra una fecha reciente sobre datos detenidos	El modelo se entrena con información vencida y produce predicciones que informan decisiones sobre una realidad que ya cambió, sin que el equipo lo advierta	Probabilidad: Alta Impacto: Mayor Zona: Alto	Preventivo. Contrastar la fecha de actualización de los datos con la frecuencia declarada antes de entrenar, según el paso 3 de la guía para entidades y la sección 10.2 para reutilizadores. Detectivo. Responsable: entidad productora y equipo técnico de datos
R2. Calidad del conjunto	Columnas sin descripción en el diccionario, vacíos en campos principales y códigos territoriales o sectoriales no estandarizados	El equipo invierte la mayor parte del tiempo en depurar y deducir el significado de las variables, y los errores no detectados se propagan al modelo	Probabilidad: Muy alta Impacto: Moderado Zona: Alto	Preventivo. Autoevaluación de aptitud para IA y umbral de nivel Preparado como condición de ingreso a la sección de IA del portal, con perfilado previo de los datos por parte del reutilizador. Detectivo. Indicadores IND-04 del Anexo E. Responsable: entidad productora

Riesgo	Causa	Consecuencia sobre el sistema de IA	Valoración inherente	Control, propósito y responsable
R3. Sesgo de representación	Los registros administrativos reflejan la capacidad institucional desigual de reporte y subrepresentan territorios y poblaciones con menor cobertura	El modelo amplifica la desigualdad del registro y produce resultados que perjudican de manera sistemática a los grupos peor documentados	Probabilidad: Media Impacto: Mayor Zona: Alto	Preventivo. Examen de representatividad antes del entrenamiento, en la fase de evaluación de CRISP-ML(Q), y registro de las decisiones adoptadas en el campo 24 de la ficha del Anexo D. Detectivo. Prueba de desempeño diferenciado por subgrupo antes del despliegue. Responsable: reutilizador
R4. Riesgo jurídico	Licencia ausente, restrictiva o incompatible con el uso previsto, o presencia de datos personales o de información con reserva legal en el conjunto publicado	El proyecto queda expuesto a la suspensión del sistema y a la responsabilidad de la entidad, con independencia de la calidad técnica del modelo	Probabilidad: Baja Impacto: Catastrófico Zona: Alto	Preventivo. Verificación de ausencia de datos personales y de información reservada antes de publicar (paso 5) y comprobación de la licencia y de sus obligaciones antes de entrenar (bloque 11 del Anexo C). Detectivo. Indicador IND-03 del Anexo E. Responsable: entidad productora y reutilizador
R5. Reidentificación de personas	El cruce de un conjunto aparentemente anónimo con otras fuentes abiertas permite reconstruir identidades que ninguna de las fuentes revela por separado	Vulneración del derecho al habeas data, con daño reputacional para la entidad publicadora y para el ecosistema de datos abiertos en su conjunto	Probabilidad: Media Impacto: Catastrófico Zona: Extremo	Preventivo. Prueba de cruce con fuentes externas y aplicación de agregación o anonimización antes de publicar, y declaración obligatoria en el campo 23 de la ficha del Anexo D como requisito de publicación del caso. Detectivo. Verificación editorial del equipo técnico. Responsable: entidad productora y reutilizador
R6. Dependencia tecnológica	El proyecto consume recursos de la capa federada, alojados en plataformas externas cuya continuidad, esquema y disponibilidad el equipo no controla	La interrupción o el cambio de estructura del servicio externo detiene el flujo de datos y obliga a reconstruir la etapa de extracción	Probabilidad: Media Impacto: Moderado Zona: Moderado	Preventivo. Preferencia por el núcleo oficial del portal con acceso programático nativo y registro del recurso federado como dependencia externa en el bloque 5 del Anexo C. Correctivo. Plan de sustitución de la fuente documentado en el bloque 14. Responsable: reutilizador

El riesgo de vigencia surge cuando el modelo se entrena con series detenidas que aparentan actualidad por el movimiento de sus metadatos, y se gestiona con la verificación de la fecha efectiva de los datos que prescriben ambas guías. En relación a la calidad, este riesgo aparece cuando se identifican columnas sin describir, vacíos y códigos no estandarizados, y se gestiona con la autoevaluación y con el perfilado previo de los datos. El riesgo de sesgo proviene de registros administrativos que subrepresentan poblaciones o territorios, y se gestiona con el examen de representatividad de la fase de evaluación de CRISP-ML(Q).

En el aspecto jurídico, el riesgo deriva de licencias ausentes o restrictivas y de la eventual presencia de datos personales, y se gestiona con la verificación de licencia y la revisión de anonimización antes del entrenamiento. Por último, el riesgo de dependencia tecnológica aparece cuando el proyecto consume recursos federados de plataformas externas cuya continuidad no controla, y se gestiona prefiriendo el núcleo oficial del portal con acceso programático nativo.

6.5. Aspectos éticos

Los principios de la Guía Ética se aplican al dato con la misma fuerza que al modelo. La entidad que publica responde por la ausencia de datos personales, por la declaración honesta del método de recolección y por la advertencia de los sesgos conocidos; el reutilizador responde por el uso dentro de la licencia, por la no reidentificación de personas y por la explicabilidad del sistema resultante. Entre estos deberes, la prevalencia de los derechos de niñas, niños y adolescentes merece atención expresa cuando se trabaja con datos de educación, salud o protección social, dominios que concentran buena parte del catálogo nacional. En caso de conflicto entre el valor analítico de un conjunto y la protección de las personas, la protección prevalece, y la decisión de publicar o de usar se documenta con sus razones.

La estrategia exige que, antes de la publicación, el administrador de datos aplique técnicas de anonimización proporcionales al riesgo de reidentificación (supresión de identificadores directos, generalización de cuasi-identificadores y agregación de variables sensibles), y que deje constancia del procedimiento aplicado en los metadatos del conjunto. Esta exigencia opera con independencia del régimen de la Ley 1581 de 2012 y no sustituye la valoración jurídica que corresponda a cada entidad.

6.6. Casos nacionales e internacionales

La evidencia nacional muestra que el ecosistema colombiano ya produce sistemas de inteligencia artificial sobre datos abiertos en salud, seguridad, educación y gestión del territorio (Tabla 21 de la sección 10.4). En el plano internacional, los países miembros de la OCDE con mejor desempeño en el índice OURdata avanzan en la misma dirección que esta estrategia, con catálogos que etiquetan la aptitud de los datos para el aprendizaje automático, portales que publican casos de uso como material de referencia y programas públicos de formación de reutilizadores (OECD, 2026). La Unión Europea ofrece el referente regulatorio más desarrollado, con la obligación de publicar sus categorías de datos de alto valor por interfaces de programación y sin costo (Comisión Europea, 2023), condición que el núcleo oficial de datos.gov.co ya cumple de manera nativa para todo su catálogo estructurado.

La segunda edición del Global Data Barometer aporta el contraste regional. La medición encuentra que la inteligencia artificial gana presencia en las estrategias nacionales de América Latina mientras pocos países disponen de políticas claras para orientar su uso ético, y que los marcos de gobernanza rara vez abordan el sesgo algorítmico, la calidad de los datos o la rendición de cuentas de las decisiones automatizadas (Global Data Barometer, 2025).

Colombia cuenta con la Guía Ética, con el CONPES 4144 y con los instrumentos de esta estrategia, lo que la sitúa en condiciones de convertir esa ventaja normativa en resultados verificables de reutilización. Las historias del portal documentan además experiencias regionales de reutilización con propósito, entre ellas el Programa Interamericano de Datos Abiertos para la Lucha contra la Corrupción y las iniciativas de datos para la acción climática, que ofrecen plantillas de trabajo replicables por las entidades del país.

7. Fortalecimiento de metadatos y catálogos

Este capítulo desarrolla el segundo componente del hito del CONPES y responde la pregunta por la articulación entre calidad, metadatos y catálogos. Los metadatos forman la capa que permite a un desarrollador decidir, antes de abrir el archivo, si un conjunto sirve para su modelo, con qué cobertura, con qué licencia y con qué frecuencia de actualización. El propósito consiste en enriquecer los metadatos, los vocabularios y los catálogos del portal con una orientación explícita hacia la inteligencia artificial, y así reducir el tiempo de preparación de datos que hoy concentra buena parte de cualquier proyecto analítico.

7.1. El esquema de metadatos actual de datos.gov.co

El portal ya dispone de una base de metadatos sólida que conviene reconocer antes de proponer cambios. La plataforma Socrata que opera bajo el portal datos.gov.co registra de forma nativa el título, la descripción, la entidad a la que se atribuye el conjunto, el correo de contacto, la licencia, las fechas de creación y de última actualización de los datos y de los metadatos, un diccionario con la descripción de cada columna y el acceso automático por interfaz de programación.

Sobre esa base, el portal añade dos grupos de campos personalizados. El grupo Información de Datos incluye el idioma, la cobertura geográfica y la frecuencia de actualización. El grupo Información de la Entidad reúne el sector, el orden nacional o territorial, el departamento, el municipio y el código DIVIPOLA de la entidad. El esquema cubre buena parte de lo que un catálogo de datos abiertos necesita y ofrece un punto de partida favorable. El diagnóstico también observa oportunidades de mejora en la aplicación del esquema, entre ellas descripciones de columnas por completar, campos de entidad que difieren de la atribución del conjunto y un diccionario de datos que se diligencia de forma parcial.

7.2. Campos mínimos de metadatos para IA

La estrategia propone un grupo de campos mínimos que todo conjunto de datos priorizado debería diligenciar, y los clasifica según su disponibilidad actual en el portal. La Tabla 12 distingue los campos que ya existen de forma nativa o personalizada, los que existen de manera parcial y conviene fortalecer, y los que se incorporarían para habilitar el uso responsable de la inteligencia artificial. Esta distinción aprovecha lo que ya funciona y concentra el esfuerzo en las adiciones que aportan valor diferencial.

Tabla 12. Campos mínimos de metadatos para IA, estado actual en datos.gov.co y acción propuesta

Campo	Descripción	Estado en datos.gov.co	Acción propuesta
Título y descripción	Nombre y resumen del conjunto	Disponible	Mantener
Entidad y responsable	Atribución y propietario del dato	Disponible	Mantener
Correo de contacto	Contacto para la gestión del dato	Disponible	Mantener. Hacer obligatoria
Licencia	Licencia que autoriza la reutilización	Disponible	Mantener. Hacer obligatoria
Idioma y cobertura geográfica	Idioma y nivel territorial	Disponible	Mantener
Frecuencia de actualización	Periodicidad declarada de publicación	Disponible	Mantener
Sector, orden y ubicación	Clasificación institucional de la entidad	Disponible	Mantener
Fechas de actualización	De datos y de metadatos	Disponible	Mantener
Formato y acceso (API)	CSV, JSON, GeoJSON y API	Disponible	Mantener
Diccionario de datos	Descripción de cada variable y su tipo	Disponible	Mantener. Hacer obligatoria
Cobertura temporal	Periodo, con inicio y fin, que abarca el conjunto	Por incorporar	Añadir campo estructurado
Nivel de sensibilidad / PII	Presencia de datos personales	Por incorporar	Añadir etiqueta obligatoria
Método de recolección / linaje	Origen y trazabilidad del dato	Por incorporar (hoy en texto libre)	Estructurar el campo
Nivel de aptitud para IA	Nivel de la autoevaluación, de Inicial a Listo para IA	Por incorporar	Añadir campo
Orientación de uso para IA	Casos de uso sugeridos y advertencias	Por incorporar	Añadir campo orientador

Fuente: elaboración propia con base en la inspección de la metadata de datos.gov.co, DCAT (World Wide Web Consortium, 2014) y la Guía de estándares de calidad e interoperabilidad de datos abiertos.

Los campos que se incorporan aparecen en la parte final de la tabla y comparten un rasgo común. Trasladan al catálogo información que hoy se conoce de manera tardía, cuando el proyecto de inteligencia artificial ya está en marcha. La cobertura temporal, la etiqueta de sensibilidad, el método de recolección, el nivel de la autoevaluación de aptitud para IA y la orientación de uso convierten la ficha del conjunto en un instrumento de decisión. Estos campos alimentan de manera directa el catálogo orientado a inteligencia artificial de la sección 7.5, que los usa como filtros y como criterios de recomendación.

7.3. Vocabularios, taxonomías e interoperabilidad semántica

La descubribilidad mejora cuando las entidades usan el mismo lenguaje para describir realidades equivalentes. El portal ya emplea el código DIVIPOLA del DANE para la ubicación territorial, tanto en los campos de entidad como en las variables de muchos conjuntos, lo que ofrece un punto de partida para la interoperabilidad. La estrategia extiende ese principio y promueve vocabularios controlados y taxonomías compartidas para las categorías temáticas, las unidades territoriales y las clasificaciones sectoriales, de modo que un modelo integre datos de varias entidades con equivalencias resueltas. La interoperabilidad semántica se apoya en identificadores persistentes que enlazan conjuntos relacionados y habilitan grafos de datos reutilizables por sistemas de inteligencia artificial.

7.4. Automatización del levantamiento de metadatos

El diligenciamiento manual de los campos nuevos puede apoyarse en la solución automatizada de rastreo y evaluación diseñada dentro del proyecto conjunto entre el MinTIC y la Universidad de Cartagena para el seguimiento de la Resolución 1519 de 2020. La incorporación de los campos de la Tabla 12 en el esquema del portal facilita su recuperación automática por esa vía, con lo cual el costo de mantenimiento del estándar disminuye para las entidades y la verificación del MinTIC gana escala.

7.5. El catálogo orientado a IA: sección del portal datos.gov.co

La sección de inteligencia artificial del portal materializa el tercer componente del hito del CONPES y convierte los metadatos enriquecidos en un producto visible para el usuario final. Ofrece un punto único de acceso a los datos aptos para inteligencia artificial, con la información y los recursos que un desarrollador necesita para pasar del dato al modelo. Funciona como una capa de descubrimiento y acompañamiento sobre el catálogo existente, aprovecha la infraestructura del portal y respeta los flujos de publicación de las entidades. La Tabla 13 describe sus componentes, la función de cada uno y el insumo de metadatos que consume.

Tabla 13. Componentes de la sección de IA del portal y metadatos que consumen

Componente	Función	Insumo de metadatos
Página de entrada	Presenta propósito, sectores priorizados y conjuntos de datos aptos para IA destacados (IA-Ready)	Nivel de la autoevaluación, sector
Catálogo de conjuntos aptos para IA (IA-ready filtrable)	Lista los conjuntos aptos con filtros aplicados	Calidad, licencia, formato, cobertura

Componente	Función	Insumo de metadatos
Fichas de caso de uso	Documentan aplicaciones con datos, método y resultados	Aptitud y uso para IA, sesgos
Datasets recomendados por sector	Sugieren conjuntos según dominio y demanda	Sector, métricas de uso
Recursos reutilizables	Ofrecen cuadernos, ejemplos de API y plantillas	Formato, acceso por API
Historias de uso	Relatan experiencias de reutilización	Entidad, casos de uso
Tablero de calidad	Muestra el nivel de la autoevaluación por sector y entidad	Nivel de aptitud para IA
Buscador semántico	Localiza conjuntos por significado	Vocabularios y taxonomías

Fuente: elaboración propia.

Un conjunto de datos ingresa a la sección cuando alcanza, al menos, el nivel Preparado de la autoevaluación, declara licencia abierta y expone acceso por interfaz de programación. Los casos de uso describen el problema, los conjuntos empleados, el enfoque metodológico, las métricas de desempeño y los límites de la solución, con atención especial a los sesgos y a la protección de datos. Una vez adoptada la presente guía, la sección podrá publicar los contenidos complementarios que el ecosistema ya produce, entre ellos los casos de la sección de usos, los cursos de la universidad corporativa, los videos de capacitación y los talleres prácticos, siguiendo el modelo de las historias temáticas existentes del portal.

La implementación requiere coordinación con las áreas técnicas del portal. En este sentido, se deberán solicitar los permisos técnicos sobre la plataforma, los lineamientos de experiencia de usuario, la descripción de la arquitectura actual, el esquema de administración de contenidos, la analítica web disponible y las condiciones propias de Socrata y de datos.gov.co. Estos insumos permiten dimensionar el desarrollo, elegir entre una integración nativa o una capa complementaria y asegurar que la sección respete los estándares de accesibilidad y seguridad de la Resolución 1519 de 2020.

8. Desarrollo de capacidades

Este capítulo desarrolla el cuarto componente del hito del CONPES y organiza la formación en dos rutas, una para las entidades públicas que publican datos y otra para los desarrolladores que los reutilizan. Ambas rutas se apoyan en la oferta vigente de la universidad corporativa del MinTIC y se alinean con la línea de fortalecimiento de capacidades del anexo técnico del proyecto, que ordena diseñar estrategias de formación para el uso estratégico de la inteligencia artificial, la analítica de datos y la toma de decisiones basadas en datos.

8.1. Ruta de formación para entidades

La ruta para entidades acompaña la ejecución de la guía y avanza en cuatro momentos, como describe la Tabla 14. Parte de la sensibilización directiva, continúa con la formación aplicada de los equipos en los cursos vigentes de la universidad corporativa, sigue con la aplicación asistida de los ocho pasos sobre los conjuntos priorizados de la propia entidad y culmina con la formación de formadores, que otorga autonomía para replicar el proceso en nuevas vigencias. La oferta actual incluye el curso Inteligencia Artificial Aplicada 2026, orientado a comprender el potencial de la inteligencia artificial para resolver problemas públicos y a desarrollar las primeras fases de un proyecto sin perfil de programador, y el curso Habilidades y Herramientas de IA Aplicada al Sector Público, dirigido al fortalecimiento de capacidades tecnológicas con gobernanza ética (MinTIC, 2026b).

Tabla 14. Ruta de formación para entidades públicas

Momento	Contenido	Resultado esperado
1. Sensibilización directiva	Valor público de los datos abiertos para IA, obligaciones vigentes y casos del portal	Decisión directiva de priorizar conjuntos y asignar responsables (pasos 1 y 2)
2. Formación aplicada	Cursos de la universidad corporativa del MinTIC en IA aplicada al sector público	Equipos con lenguaje común y comprensión del ciclo de un proyecto de IA
3. Aplicación asistida de la guía	Ejecución de los ocho pasos sobre los conjuntos priorizados	Conjuntos en nivel Preparado o Listo para IA registrados en el catálogo
4. Formación de formadores	Preparación de multiplicadores internos con los materiales de esta estrategia	Autonomía institucional para sostener el proceso en nuevas vigencias

Fuente: elaboración propia, alineada con la línea de fortalecimiento de capacidades del anexo técnico del proyecto.

8.2. Ruta de formación para desarrolladores

El programa para desarrolladores conecta la oferta de datos con su reutilización efectiva y se dirige a desarrolladores, científicos de datos y analistas de las entidades públicas,

a estudiantes y docentes de programas de ingeniería y afines, y a la comunidad de reutilizadores externos. El enfoque combina fundamentos conceptuales con práctica sobre conjuntos reales del portal, de modo que cada participante trabaje con datos de su propio dominio. El aprendizaje se organiza alrededor de un proyecto que recorre el ciclo completo de CRISP-ML(Q), desde la selección del conjunto hasta la publicación de resultados, y ancla los conceptos en una aplicación concreta.

El programa se organiza en siete temas que reproducen el ciclo de vida del dato en un proyecto de inteligencia artificial. Cada tema reúne sus contenidos y la razón que justifica su inclusión, y su alcance se ajusta al público y al nivel de cada cohorte, como presenta la Tabla 15. La duración, la profundidad y el calendario se precisan en la fase de implementación, según lo defina el anexo técnico correspondiente.

Tabla 15. Temas del programa de capacitación para desarrolladores, contenidos y justificación

Temas	Contenidos	Justificación de su inclusión
Fundamentos y marco normativo	Datos abiertos, CONPES 4144, Resolución 1519 y ética pública	Sitúa el programa en el marco de política y crea un lenguaje común entre entidades y desarrolladores
Calidad y preparación de datos	Perfilado, tratamiento de vacíos y atípicos, y uso de la autoevaluación de aptitud para IA	Lleva el eje E1 a la práctica, porque la calidad de los datos define el desempeño de los modelos
Metadatos e interoperabilidad	Estándar de metadatos, DCAT, vocabularios controlados y catálogos	Desarrolla el eje E2 y logra que los datos se encuentren y se comprendan sin fricción
Acceso programático a los datos	Consumo de las API de datos.gov.co y Socrata, y automatización de la extracción	Reduce la barrera de acceso y habilita flujos de datos reproducibles hacia los modelos
Ética, sesgos y protección de datos	Sesgos, datos personales, transparencia y explicabilidad de los modelos	Protege los derechos y la confianza pública en las aplicaciones de IA del Estado
Del dato al modelo	Preparación, entrenamiento, validación e interpretación de resultados	Conecta la oferta de datos con soluciones concretas y consolida la competencia central del eje E3
Publicación y reutilización de resultados	Documentación, fichas de caso y publicación en la sección de IA del portal	Cierra el ciclo, alimenta la comunidad y genera recursos reutilizables para otras entidades

Fuente: elaboración propia.

El programa privilegia la práctica sobre datos reales y se apoya en materiales abiertos y reutilizables, que incluyen guías, cuadernos de análisis y conjuntos de práctica extraídos del portal. El aprendizaje se valida con productos aplicados antes que con pruebas teóricas, y el proyecto integrador culmina en una ficha de caso de uso susceptible de publicarse en la sección de inteligencia artificial del portal, con lo que la formación alimenta de manera directa el catálogo de casos.

8.3. Certificación, comunidad e indicadores

Se recomienda que el programa otorgue una certificación por competencias a quienes completen el proyecto integrador, y que establezca una comunidad de práctica que sostenga el intercambio entre cohortes. Esa comunidad ofrece un espacio para compartir cuadernos, resolver dudas técnicas y difundir nuevos conjuntos aptos, y alimenta los indicadores de reutilización, casos de uso publicados y desarrolladores certificados que el paso 8 de la guía para entidades contempla.

La sostenibilidad de ambas rutas descansa en condiciones, que son: la institucionalización de los instrumentos dentro de los procesos regulares del ministerio y de las entidades, la formación de capacidades internas mediante el esquema de formación de formadores y la retroalimentación continua entre la comunidad de reutilizadores y las entidades productoras, que orienta la priorización hacia los datos que generan valor.

9. Guía para la implementación, desarrollo y uso de sistemas de inteligencia artificial en entidades públicas de colombia

Este capítulo responde la primera pregunta que la estrategia se propone resolver, esto es, cómo debe preparar una entidad pública sus datos con el fin de que resulten útiles a los sistemas de inteligencia artificial. La respuesta se organiza en ocho pasos que recorren el ciclo completo del dato, desde su identificación hasta el seguimiento de su uso. Cada paso indica qué hacer, quién lo hace, con qué instrumento y cómo verificar que quedó bien hecho. Los pasos se apoyan en obligaciones ya vigentes y en la infraestructura del portal, de modo que ninguno crea cargas nuevas para la entidad (Figura 3). Una entidad que ejecute los ocho pasos sobre sus conjuntos priorizados tendrá datos en condición de alimentar modelos de inteligencia artificial y podrá demostrarlo con la autoevaluación registrada en el catálogo.



Figura 3. Guía ética para la implementación, desarrollo y uso de sistemas de inteligencia artificial en entidades públicas de colombia

Fuente: elaboración propia

9.1. Paso 1. Identificar los datos de alto valor

El punto de partida es el inventario de activos de información que la entidad ya mantiene por mandato de la Ley 1712 de 2014. Sobre ese inventario, el equipo directivo y el responsable de datos identifican los conjuntos con potencial para inteligencia artificial aplicando tres preguntas:

- ¿Existe demanda comprobada, verificable en las métricas de consulta del portal, en las peticiones ciudadanas y en las solicitudes de otras entidades?
- ¿El conjunto sostiene decisiones sensibles o servicios esenciales, como ocurre en salud, seguridad o inclusión social?
- ¿El conjunto habilitaría casos de uso emergentes de alto retorno, como los datos ambientales para alertas tempranas o los agropecuarios para la gestión del riesgo climático?

La clasificación resultante en datos de alto valor, datos críticos y datos estratégicos, ordena todo el trabajo posterior.

9.2. Paso 2. Priorizar

Pocas entidades pueden llevar todos sus conjuntos a la condición de listos para inteligencia artificial en un mismo ejercicio, de modo que conviene concentrar el esfuerzo donde el retorno es mayor. La priorización combina la demanda observada, la relevancia sectorial y el juicio experto de la propia entidad, y produce una lista corta de conjuntos sobre los cuales se ejecutan los pasos siguientes. Como referencia, la Tabla 9 relaciona los sectores con mayor volumen de datos en el portal con los casos de uso de inteligencia artificial de mayor potencial, y puede usarse como criterio de contraste. El resultado del paso es un acta breve de priorización con los conjuntos elegidos, el responsable de cada uno y la fecha objetivo de mejora.

9.3. Paso 3. Asegurar la calidad con la autoevaluación de aptitud para IA

La calidad se gestiona con la autoevaluación de aptitud para inteligencia artificial, el instrumento que desarrolla el primer componente del hito del CONPES. La persona responsable del conjunto revisa ocho aspectos y marca los que se cumplen. Cada aspecto se describe en palabras simples y remite a algo observable en el archivo o en la ficha, sin fórmulas ni ponderaciones, de manera que cualquier persona clasifique su conjunto en pocos minutos sin conocimientos de estadística ni de programación. La Tabla 16 presenta la lista de verificación.

Tabla 16. Lista de verificación para la autoevaluación de aptitud para IA

Aspecto	Qué comprobar (en palabras simples)
Estructura clara	El archivo abre en una hoja de cálculo y cada columna tiene un título entendible en la primera fila
Datos completos	Las columnas principales tienen valor en casi todas las filas, sin grandes vacíos
Datos al día	La información corresponde a la fecha más reciente que promete la ficha del conjunto
Diccionario de datos	Cada columna cuenta con una descripción breve que explica qué contiene
Licencia de uso	La ficha muestra de forma visible la licencia que permite reutilizar los datos
Acceso fácil	El conjunto se descarga o se consulta por un enlace estable o una interfaz de programación
Códigos estándar	Los municipios, departamentos y categorías usan los códigos oficiales, como la DIVIPOLA del DANE
Origen del dato	La ficha explica de dónde provienen los datos y cómo se recogen

Fuente: elaboración propia. La persona marca los aspectos que su conjunto cumple.

El número de aspectos marcados produce el valor del Índice de Aptitud para Inteligencia Artificial, que se obtiene al dividir los aspectos cumplidos entre ocho y multiplicar por cien. Un conjunto que cumple seis de los ocho aspectos obtiene un IAA de 75. El valor se expresa en una escala de cero a cien y equivale al nivel de la Tabla 17 según las bandas que allí se describen, con lo cual la entidad dispone de una cifra comparable y de una etiqueta comprensible sobre el mismo trabajo de verificación. El promedio de los valores de los conjuntos priorizados de una entidad o de un sector produce el índice de calidad agregado.

Con los aspectos marcados, la persona elige el nivel cuya descripción se parece más a su conjunto. Los niveles van de Inicial a Listo para IA y cada uno indica en palabras claras lo que el conjunto debe cumplir, como se describe en la Tabla 17. Se selecciona el nivel más alto cuya descripción se cumpla por completo y, ante la duda, el inmediatamente inferior. El nivel Listo para IA se reserva para los conjuntos que cumplen los ocho aspectos de la lista.

Tabla 17. Niveles de la autoevaluación de aptitud para IA

Valor del IAA	Aspectos cumplidos	Nivel	Qué debe cumplir el conjunto
0 a 49	0 a 3 de 8	Inicial	Está publicado y se puede abrir, aunque presenta vacíos, columnas sin describir o sin licencia clara. Cumple pocos aspectos de la lista
50 a 74	4 a 5 de 8	En desarrollo	Tiene estructura clara y datos completos, y avanza en el diccionario y en la licencia.

Valor del IAA	Aspectos cumplidos	Nivel	Qué debe cumplir el conjunto
			Cumple alrededor de la mitad de los aspectos
75 a 99	6 a 7 de 8	Preparado	Cumple casi todos los aspectos, con diccionario, licencia y acceso fácil, y solo quedan detalles menores por ajustar
100	8 de 8	Listo para IA	Cumple los ocho aspectos, con datos al día, licencia abierta declarada, diccionario completo, acceso por enlace o interfaz de programación y códigos estándar

Fuente: elaboración propia.

La autoevaluación señala el camino de mejora, donde un conjunto en nivel Inicial avanza cuando ordena sus columnas y completa los vacíos. Uno en nivel En desarrollo progresa cuando agrega la descripción de cada columna y declara la licencia. Uno en nivel Preparado alcanza la condición de Listo para IA cuando asegura el acceso por enlace o interfaz de programación y adopta los códigos estándar. Cada movimiento es concreto y está al alcance de la entidad productora sin apoyo técnico especializado. El nivel elegido se guarda como un campo en la ficha del conjunto y viaja con él en el catálogo; se recomienda que el MinTIC consolide los niveles en un tablero de avance y verifique una muestra para confirmar que la autoevaluación refleja el estado real.

9.4. Paso 4. Documentar los metadatos

Los metadatos permiten que un reutilizador decida, antes de abrir el archivo, si el conjunto sirve para su modelo. En este paso la entidad diligencia los campos mínimos del estándar, con atención particular al diccionario de datos con la descripción de cada columna, a la licencia, a la cobertura temporal y a la etiqueta de presencia de datos personales. La experiencia del diagnóstico muestra que el diccionario incompleto es la carencia más frecuente del catálogo; conjuntos de consulta masiva, como el código único de medicamentos, mantienen la mayoría de sus columnas sin descripción, lo que obliga a cada reutilizador a deducir el significado de las variables. Completar el diccionario es la intervención de mejor relación entre esfuerzo y beneficio de toda la guía.

9.5. Paso 5. Publicar

La publicación se realiza en el Portal Nacional de Datos Abiertos conforme al anexo 2 de la Resolución 1519 de 2020. La entidad carga el conjunto en formato estructurado, con codificación UTF-8, títulos de columna en la primera fila y una fila por registro, y verifica que el portal exponga el acceso por interfaz de programación, que la plataforma genera de forma automática. La licencia abierta se declara en el campo correspondiente de la

ficha; la declaración explícita despeja el riesgo jurídico del entrenamiento de modelos y por esa razón la estrategia la convierte en requisito de la priorización. Antes de publicar, el responsable confirma que el conjunto no contiene datos personales ni información clasificada o reservada, con el procedimiento de anonimización o agregación que corresponda.

9.6. Paso 6. Actualizar

La desactualización de los datos fuente degrada de forma directa la validez de las predicciones del modelo, por lo que la actualización sostiene el valor de todo lo anterior. La entidad declara una frecuencia de actualización realista y la cumple, y revisa de forma periódica la diferencia entre la fecha de actualización de los datos y la de los metadatos, porque la segunda suele moverse sin la primera y genera una apariencia de vigencia que el diagnóstico documentó en conjuntos de alta consulta. Cuando un conjunto deja de actualizarse por cambios institucionales o de fuente, la ficha lo declara de manera explícita para que los reutilizadores conozcan el corte real de la serie.

9.7. Paso 7. Gobernar

La gobernanza asigna responsables y periodicidad donde los pasos anteriores operen como proceso y no como campaña. La coordinación general de la estrategia recae en la Dirección de Gobierno Digital del MinTIC, cada entidad productora asume la calidad y la actualización de sus conjuntos, y un equipo técnico administra la autoevaluación, el catálogo orientado a inteligencia artificial y las rutas de formación. La Tabla 18 describe los roles, su responsabilidad y la periodicidad con que actúan. Dentro de la entidad, conviene que el comité institucional de gestión y desempeño incorpore el avance de la autoevaluación entre sus temas de revisión, de modo que la apertura de datos cuente con respaldo directivo permanente.

Tabla 18. Roles, responsabilidades y periodicidad del modelo de gobernanza

Rol	Responsabilidad	Periodicidad
Coordinación (MinTIC)	Dirige la estrategia, aprueba lineamientos y consolida indicadores	Permanente
Entidad productora	Publica, actualiza y mejora la calidad de sus conjuntos	Según frecuencia declarada
Equipo técnico de datos	Verifica la autoevaluación, gestiona el catálogo y la sección de IA	Anual

Fuente: elaboración propia.

9.8. Paso 8. Hacer seguimiento

El seguimiento cierra el ciclo con un conjunto acotado de indicadores que cubren la apertura, la calidad y el uso. Los indicadores se calculan sobre los conjuntos priorizados y se publican en el tablero de seguimiento, lo que permite a las entidades comparar su avance y rendir cuentas. La Tabla 19 presenta los indicadores, su fórmula, la frecuencia de medición y el responsable. Cabe resaltar que cada indicador se desarrolla en el Anexo E mediante una ficha técnica que precisa su definición, su universo de medición, su línea base, su meta propuesta, sus desagregaciones y sus limitaciones de interpretación, de manera que el cálculo resulte reproducible y auditable por el equipo responsable del tablero de seguimiento.

Tabla 19. Indicadores de seguimiento de la estrategia

Indicador	Fórmula	Frecuencia	Responsable
Conjuntos listos para IA	N.º en nivel Listo para IA / total priorizado	Anual	Equipo técnico
Conjuntos preparados o superiores	N.º en nivel Preparado o Listo para IA / total priorizado	Anual	Equipo técnico
Licencia abierta	N.º de conjuntos con licencia abierta / total priorizado	Anual	Entidad productora
Metadatos completos	N.º con campos mínimos diligenciados / total priorizado	Anual	Entidad productora
Reutilización	N.º de reutilizaciones y descargas registradas por periodo	Anual	Coordinación (MinTIC)
Casos de uso publicados	N.º de fichas de caso de IA publicadas en el portal	Anual	Coordinación (MinTIC)
Desarrolladores certificados	N.º de participantes certificados por el programa	Anual	Coordinación (MinTIC)

Fuente: elaboración propia.

El recorrido de los ocho pasos exige capacidades que no todas las entidades tienen instaladas. Para cerrar esa brecha, los servidores públicos disponen del curso Inteligencia Artificial Aplicada 2026 de la universidad corporativa del MinTIC, que capacita en la comprensión del potencial de la inteligencia artificial para resolver problemas públicos y en el desarrollo de las primeras fases de un proyecto sin requerir perfil de programador (MinTIC, 2026b). El capítulo 8 resume la ruta de formación para entidades, que articula este curso con la aplicación práctica de la presente guía.

10. Guía para reutilizadores

Este capítulo se dirige a quienes convierten los datos abiertos en soluciones, es decir, a la academia, los investigadores, los desarrolladores, las empresas, los emprendedores y la sociedad civil. Responde a dos de las preguntas orientadoras de la estrategia, la que indaga por las buenas prácticas que deben seguir los desarrolladores al utilizar datos abiertos y la que examina qué características de los datos facilitan su reutilización. La guía organiza el recorrido alrededor de una metodología de referencia (Figura 4), muestra dónde y cómo encontrar los datos, explica el acceso programático al portal, presenta casos nacionales que demuestran lo que ya es posible y cierra con los criterios de desarrollo responsable que la Guía Ética exige a todo sistema de inteligencia artificial construido sobre información pública.



Figura 4. Guía para reutilizadores

Fuente: elaboración propia

10.1. La metodología CRISP-ML como ruta de trabajo

La estrategia adopta como referencia la metodología CRISP-ML(Q), un modelo de proceso para proyectos de aprendizaje automático con aseguramiento de calidad en

cada fase (Studer et al., 2021), que el curso de Inteligencia Artificial Aplicada de la universidad corporativa del MinTIC propone como estructura de trabajo (MinTIC, 2026b). Su valor para el reutilizador de datos abiertos está en que ordena las decisiones y obliga a verificar la viabilidad antes de invertir en el modelado. Buena parte de los proyectos que fracasan lo hacen en las primeras fases, cuando el equipo descubre tarde que los datos disponibles no alcanzan para el objetivo propuesto, por ende, la metodología convierte ese descubrimiento en una verificación temprana y económica. La Tabla 20 describe las fases y las traduce al trabajo con datos abiertos del portal.

Tabla 20. Fases de CRISP-ML(Q) aplicadas a proyectos con datos abiertos

Fase	Actividades principales	Aplicación con datos abiertos de datos.gov.co
Entendimiento del negocio y de los datos	Definir el problema público, los criterios de éxito y la viabilidad; inventariar y evaluar las fuentes	Formular el problema en términos de decisión pública; buscar conjuntos en el portal y revisar nivel de aptitud, licencia, cobertura y vigencia antes de comprometerse
Preparación de los datos	Limpiar, integrar, transformar y documentar los datos	Validar la vigencia efectiva contra la fecha de actualización de los datos; resolver equivalencias con códigos DIVIPOLA; registrar cada transformación
Modelado	Seleccionar, entrenar y ajustar los modelos	Documentar qué conjuntos y qué cortes alimentan cada experimento, para que el resultado sea reproducible
Evaluación	Medir el desempeño contra los criterios de éxito y revisar sesgos	Contrastar el modelo con los sesgos conocidos declarados en los metadatos y con poblaciones subrepresentadas
Despliegue	Poner el modelo en operación	Declarar la fuente y la licencia de los datos; citar los conjuntos con su enlace persistente
Monitoreo y mantenimiento	Vigilar la degradación del desempeño y actualizar	Monitorear la actualización de los conjuntos fuente; reentrenar cuando la serie avance o cambie de estructura

Fuente: elaboración propia a partir de Studer et al. (2021) y del curso Inteligencia Artificial Aplicada 2026 (MinTIC, 2026b).

El instrumento complementario de esta fase inicial es el Machine Learning Canvas, incluido como anexo C, que permite evaluar en una sola página la viabilidad de un proyecto de inteligencia artificial considerando la disponibilidad de datos, las restricciones normativas y los requisitos de robustez, escalabilidad, explicabilidad y uso eficiente de recursos. Se recomienda diligenciar antes de iniciar cualquier desarrollo, para así evitar inversiones sobre proyectos inviables.

10.2. Cómo encontrar los datos

El punto de entrada es el buscador del Portal Nacional de Datos Abiertos, disponible en datos.gov.co, que permite filtrar por categoría temática, entidad, etiqueta y tipo de activo. La ficha de cada conjunto concentra la información de decisión, es decir, la descripción, la entidad responsable, la licencia, la frecuencia declarada, las fechas de actualización de los datos y de los metadatos, y el diccionario de columnas. La lectura de la ficha ahorra tiempo, destacando dos verificaciones que merecen mención:

- Comparar la fecha de actualización de los datos con la frecuencia declarada, porque los metadatos pueden moverse sin que la serie avance.
- Confirmar la licencia, que autoriza la reutilización y define las condiciones de atribución.

Cuando la sección de inteligencia artificial del portal entre en operación con el catálogo orientado a IA, el reutilizador podrá filtrar además por nivel de aptitud, cobertura temporal y presencia de datos personales.

10.3. Acceso mediante interfaces de programación

Cada conjunto oficial del portal se expone de forma automática mediante la interfaz de programación de la plataforma Socrata, sin registro previo para consultas básicas. El patrón de acceso es estable, ya que cada conjunto tiene un identificador de nueve caracteres visible en su dirección web, y el recurso se consume añadiendo ese identificador al punto de acceso del portal, con la forma www.datos.gov.co/resource/identificador.json. La interfaz admite parámetros de consulta del lenguaje SoQL, entre ellos la selección de columnas, los filtros por valor, el ordenamiento y la paginación, lo que permite extraer solo los registros necesarios en lugar de descargar el conjunto completo. Para cargas de trabajo intensivas conviene solicitar un token de aplicación, que amplía los límites de consulta. Los formatos de salida incluyen JSON, CSV y GeoJSON para los conjuntos con geometría, con lo cual el flujo de extracción se integra a cualquier lenguaje de análisis sin pasos intermedios manuales. La documentación completa de la interfaz, con sus parámetros, límites y ejemplos por lenguaje, se publica en el portal para desarrolladores de la plataforma (Tyler Technologies, s.f.). Esta vía programática es la que hace reproducibles los proyectos, porque el cuaderno de análisis que consume la interfaz puede ejecutarse de nuevo cuando la serie se actualice.

10.4. Casos de uso como material de aprendizaje

La sección de usos del portal, disponible en herramientas.datos.gov.co/usos, reúne los productos que ciudadanos, universidades, empresas y entidades han construido con datos abiertos, con filtros por tipo de producto, departamento y sector, y constituye un material de estudio para un reutilizador que inicia. Los casos recientes muestran una presencia creciente de la inteligencia artificial. La Tabla 21 selecciona cinco casos publicados que ilustran enfoques distintos, desde asistentes conversacionales hasta modelos predictivos, y que pueden consultarse en la misma sección.

Tabla 21. Casos nacionales de reutilización con inteligencia artificial publicados en la sección de usos

Caso	Descripción	Datos y enfoque
MIA, Motor de Inteligencia Académica	Plataforma de búsqueda que interpreta consultas en lenguaje natural sobre programas de educación superior	Datos del SNIES y de datos.gov.co integrados con información geográfica; asistente con IA
AlertaMaterna	Sistema de clasificación de riesgo obstétrico y predicción de mortalidad infantil en la Orinoquía	Datos abiertos de salud; modelo de clasificación y predicción
CrimeLab Santander	Plataforma que integra y analiza información oficial sobre criminalidad departamental	Datos de seguridad y defensa; visualización y predicción
SynergIA	Ecosistema de seguridad ciudadana que combina IA con análisis geoespacial predictivo	Datos territoriales; analítica geoespacial predictiva
DACC	Agente inteligente que ayuda a las entidades a optimizar el uso de sus activos de datos abiertos	Metadatos del catálogo; agente conversacional

Fuente: elaboración propia a partir de la sección de usos del portal (herramientas.datos.gov.co/usos, consulta de julio de 2026).

Las historias temáticas del portal complementan estos casos con recorridos guiados, entre ellas la del Programa Interamericano de Datos Abiertos para la Lucha contra la Corrupción, las dedicadas a la acción climática y la guía introductoria para quienes usan datos abiertos por primera vez. Se recomienda a los reutilizadores postular sus propios productos en la misma sección, porque cada caso publicado amplía el material de aprendizaje disponible y da visibilidad al autor ante el ecosistema.

10.5. Desarrollo responsable de sistemas de inteligencia artificial

El uso de datos públicos impone deberes proporcionales a su alcance. El reutilizador que construye sistemas de inteligencia artificial sobre datos abiertos aplica los principios de la Guía Ética en su propio ámbito, lo que en la práctica se concreta en cuatro conductas:

- Verificar que los datos no permitan reidentificar personas, incluso cuando el conjunto publicado parezca anónimo, porque el cruce con otras fuentes puede revelar identidades.
- Examinar los sesgos de representación de los datos antes de entrenar, dado que las series administrativas reflejan las desigualdades del registro y un modelo las amplifica si no se corrigen.
- Documentar el sistema de modo que sus resultados puedan explicarse ante los afectados, condición que la guía denomina transparencia y explicación.
- Mantener el control humano de las decisiones cuando el sistema apoye actuaciones públicas o afecte derechos.

Estas conductas no son opcionales para quien aspire a publicar su caso en el portal o a proveer soluciones al Estado, porque los criterios de la ficha de caso del anexo D las incorporan de manera expresa.

10.6. Buenas prácticas del reutilizador

La experiencia acumulada en la sección de usos y en los concursos de datos permite condensar las prácticas que distinguen a los proyectos sostenibles, donde el reutilizador que sigue esta guía:

- Valida la vigencia efectiva de cada conjunto antes de usarlo.
- Consume los datos por la interfaz de programación para que su flujo sea reproducible.
- Registra las transformaciones que aplica.
- Cita los conjuntos con su enlace y su entidad responsable.
- Respeta las condiciones de la licencia declarada.
- Publica su ficha de caso cuando el producto madura.
- Reporta a la entidad productora los errores de calidad que detecta.

Referencias

Comisión Europea. (2023). Reglamento de Ejecución (UE) 2023/138 de la Comisión, por el que se establece una lista de conjuntos de datos específicos de alto valor y las modalidades de su publicación y reutilización. Diario Oficial de la Unión Europea. https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj

Congreso de la República de Colombia. (2012). Ley 1581 de 2012, por la cual se dictan disposiciones generales para la protección de datos personales. Bogotá, Colombia.

Congreso de la República de Colombia. (2014). Ley 1712 de 2014, Ley de Transparencia y del Derecho de Acceso a la Información Pública Nacional. Bogotá, Colombia.

Departamento Administrativo de la Función Pública. (2022). Guía para la administración del riesgo y el diseño de controles en entidades públicas (versión 6). Bogotá, Colombia: DAFP.

Departamento Nacional de Planeación. (2025). Documento CONPES 4144: Política Nacional de Inteligencia Artificial. Bogotá, Colombia: DNP.

Dorard, L. (2015). The Machine Learning Canvas. https://www.ownml.co/machine-learning-canvas

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H. y Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

Global Data Barometer. (2025). Global Data Barometer: Second edition. https://globaldatabarometer.org

Gobierno de Colombia. (2025). Guía Ética para la Implementación, Desarrollo y Uso de Sistemas de Inteligencia Artificial en Entidades Públicas de Colombia. Bogotá, Colombia. https://www.mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/425888:Guia-Etica-para-la-Implementacion-Desarrollo-y-Uso-de-Sistemas-de-Inteligencia-Artificial-en-Entidades-Publicas-de-Colombia

Janssen, M., Charalabidis, Y. y Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258-268.

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.-a). Guía para el uso y aprovechamiento de Datos Abiertos en Colombia. Bogotá, Colombia: MinTIC.

Ministerio de Tecnologías de la Información y las Comunicaciones. (s.f.-b). Guía de estándares de calidad e interoperabilidad de datos abiertos. Bogotá, Colombia: MinTIC.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2020). Resolución 1519 de 2020, por la cual se definen los estándares y directrices para publicar la información señalada en la Ley 1712 de 2014. Bogotá, Colombia: MinTIC.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2022). Decreto 767 de 2022, por el cual se establecen los lineamientos generales de la Política de Gobierno Digital y se subroga el capítulo 1 del título 9 de la parte 2 del libro 2 del Decreto 1078 de 2015. Bogotá, Colombia.

Ministerio de Tecnologías de la Información y las Comunicaciones. (2026a). Reportes del Portal Nacional de Datos Abiertos [conjunto de tableros]. https://www.datos.gov.co/stories/s/pvyw-9yqs

Ministerio de Tecnologías de la Información y las Comunicaciones. (2026b). Inteligencia Artificial Aplicada 2026 [curso de la universidad corporativa APTI]. https://lms.mintic.gov.co/course/view.php?id=382

Ministerio de Tecnologías de la Información y las Comunicaciones, Departamento Nacional de Planeación y Departamento Administrativo de la Presidencia de la República. (2022). Plan Nacional de Infraestructura de Datos: documento técnico y hoja de ruta. Bogotá, Colombia. https://infraestructuradatos.gov.co

OECD. (2026). Digital Government Outlook 2026: Colombia. OCDE Publishing. https://www.oecd.org/en/publications/digital-government-outlook-2026_d46c0555-en/colombia_cc2c0193-en.html

Open Data Charter. (2015). International Open Data Charter: Principles. https://opendatacharter.net

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S. y Müller, K. R. (2021). Towards CRISP-ML(Q): A machine learning process model with

quality assurance methodology. Machine Learning and Knowledge Extraction, 3(2), 392-413. https://doi.org/10.3390/make3020020

Tyler Technologies. (s.f.). Socrata Open Data API: Developer documentation. https://dev.socrata.com

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

World Wide Web Consortium. (2014). Data Catalog Vocabulary (DCAT). https://www.w3.org/TR/vocab-dcat/

Zuiderwijk, A. y Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. Government Information Quarterly, 31(1), 17-29 https://doi.org/10.1016/j.giq.2013.04.003

Anexo A. Evidencias

Este anexo documenta la evidencia que sustenta los aspectos por fortalecer del diagnóstico (sección 4.3, Tabla 5). Cada aspecto se ilustra con al menos tres conjuntos de datos vigentes del Portal Nacional de Datos Abiertos, con la entidad responsable, la evidencia observada y el enlace directo al conjunto. Las cifras y fechas corresponden a la consulta del 1 de julio de 2026.

A1. Vigencia de los datos

Algunos conjuntos muestran metadatos recientes junto a una última actualización de datos anterior a la frecuencia que declaran. La comparación entre la fecha de datos y la fecha de metadatos permite estimar la vigencia real antes de usar el conjunto para entrenamiento (Tabla A1).

Tabla A1. Evidencia de vigencia de los datos

Conjunto de datos	Entidad	Evidencia	Enlace
Beneficiarios Más Familias en Acción	Prosperidad Social	Frecuencia declarada trimestral, con datos a 31/07/2024 y metadatos a 18/05/2026	xfif-myr2
Tabla de retención documental	MinEducación	Frecuencia declarada anual, con datos a 30/03/2021 y metadatos a 18/05/2026	id2r-y7wh
Banco de Información Petrolera: Pozos	Servicio Geológico Colombiano	Datos a octubre de 2022 con metadatos a septiembre de 2025	4dai-7crq

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A2. Legibilidad por máquina

La capa federada remite a servicios externos y no expone un esquema tabular nativo en el portal, de modo que su lectura pasa por plataformas como ArcGIS o repositorios municipales. Estos conjuntos requieren una etapa de extracción y estructuración adicional antes del análisis (Tabla A2).

Tabla A2. Evidencia de legibilidad por máquina

Conjunto de datos	Entidad	Evidencia	Enlace
Resguardo Indígena Formalizado	Agencia Nacional de Tierras	Recurso federado a ArcGIS de la ANT, sin esquema tabular en el portal	f6du-dwd8

Conjunto de datos	Entidad	Evidencia	Enlace
Tarifas para Servicios de Energía - EPM	Alcaldía de Medellín y EPM	Recurso federado a medata.gov.co mediante archivo CSV externo	se7p-ytdd
Código postal para Bogotá D.C.	IDECA - Catastro Bogotá	Recurso federado a servicios GIS externos (REST, WMS y WFS)	wrs4-irpf

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A3. Consistencia de metadatos

Se observan conjuntos con columnas sin descripción y casos donde el campo de entidad difiere de la atribución declarada. La depuración de la descripción y la atribución mejora la trazabilidad y la selección de datos (Tabla A3).

Tabla A3. Evidencia de consistencia de metadatos

Conjunto de datos	Entidad	Evidencia	Enlace
Código Único de Medicamentos Vigentes	INVIMA	28 de 29 columnas sin descripción en el diccionario de datos	i7cb-raxc
SECOP Integrado	ANCP - Colombia Compra Eficiente	Atribución a la Alcaldía de Somondoco frente a la entidad ANCP-CCE	rpmr-utcd
Tabla de retención documental	Notaría Única de Dosquebradas	20 de 20 columnas sin descripción en el diccionario de datos	evet-w57c

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A4. Uniformidad del licenciamiento

La licencia es un campo nativo y los conjuntos revisados la declaran, con variantes entre Creative Commons BY y BY-SA según la entidad. La unificación del criterio facilita la reutilización en el entrenamiento de modelos (Tabla A4).

Tabla A4. Evidencia de licenciamiento

Conjunto de datos	Entidad	Evidencia	Enlace
SECOP II - Procesos de Contratación	ANCP - Colombia Compra Eficiente	Licencia declarada Creative Commons BY-SA 4.0	p6dx-8zbt
Tabla de retención documental	Empresas Varias de Medellín	Licencia declarada Creative Commons BY-SA 4.0	a93w-jcqe
Código Único de Medicamentos Vencidos	INVIMA	Licencia BY-SA 4.0, frente a variantes BY 4.0 en la capa federada	vwwf-4ftk

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A5. Foco analítico del catálogo

Una parte del catálogo corresponde a mosaicos cartográficos y a productos de cumplimiento documental, valiosos para la transparencia y la referencia territorial. Su identificación ayuda a orientar la priorización hacia el núcleo analítico (Tabla A5).

Tabla A5. Evidencia de foco analítico del catálogo

Conjunto de datos	Entidad	Evidencia	Enlace
Atlas Geológico de Colombia 2020: Fallas Geológicas	Servicio Geológico Colombiano	Mosaico cartográfico federado mediante geoservicio	3itm-5dx4
Inventario de movimientos en masa	Servicio Geológico Colombiano	Capa cartográfica federada de geoamenazas	nht7-gfb8
Tablas de Retención Documental	ICETEX	Producto de cumplimiento documental de la Ley 1712 de 2014	efxg-gmwq

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A6. Curaduría de activos comunitarios

La comunidad genera vistas derivadas de conjuntos oficiales, con amplia consulta y sin control de calidad institucional. Una validación previa habilita su uso confiable en proyectos de inteligencia artificial (Tabla A6).

Tabla A6. Evidencia de activos comunitarios

Conjunto de datos	Entidad	Evidencia	Enlace
LISTADO COLEGIOS BOGOTA	Vista de usuario sobre datos de MinEducación	Vista de la comunidad con 290.894 visualizaciones	qijw-htwa
MEDICAMENTOS POS	Vista de usuario sobre datos del INVIMA	Vista de la comunidad con 162.330 visualizaciones	a7iv-sme8
Colfondos	Vista de usuario sobre datos de la Superintendencia Financiera de Colombia	Vista de la comunidad con 160.190 visualizaciones	nug7-ugd3

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

A7. Ampliación de la reutilización

Un grupo reducido de conjuntos concentra la mayor parte de las visualizaciones y descargas del portal. Estos casos confirman la demanda existente y marcan la ruta para activar datos de sectores con menor uso (Tabla A7).

Tabla A7. Evidencia de concentración de la reutilización

Conjunto de datos	Entidad	Evidencia	Enlace
SECOP II - Contratos Electrónicos	ANCP - Colombia Compra Eficiente	1.736.841 visualizaciones y 208.883 descargas acumuladas	jbjy-vk9h
Código Único de Medicamentos Vigentes	INVIMA	1.546.789 visualizaciones acumuladas	i7cb-raxc
Beneficiarios Más Familias en Acción	Prosperidad Social	1.498.093 visualizaciones acumuladas	xfif-myr2

Fuente: elaboración propia con datos y metadata del catálogo de datos.gov.co (consulta del 1 de julio de 2026).

Anexo B. Matriz de trazabilidad

B.1. Trazabilidad contractual con el anexo técnico

Los cuatro componentes del Entregable 8 se descomponen en dieciséis obligaciones verificables, a las que la Tabla B1 añade dos obligaciones transversales de gobernanza y seguimiento. Cada obligación remite a un producto identificable, sea una tabla, un procedimiento, un criterio de decisión o un instrumento anexo.

Tabla B1. Trazabilidad de los componentes del anexo técnico del Entregable 8

No.	Obligación derivada del anexo técnico	Ubicación en el E8	Producto o instrumento que la satisface	Medio de verificación
1	Definir qué significa que un conjunto de datos esté listo para inteligencia artificial y fijar criterios mínimos	Capítulo 6, sección 6.1	Definición operativa de dato listo para IA y ocho criterios mínimos exigibles de manera simultánea	Lectura de la sección 6.1
2	Construir un índice de calidad aplicable por la propia entidad, sin dependencia de perfiles técnicos	Capítulo 9, sección 9.3	Lista de verificación de ocho aspectos observables (Tabla 16)	Tabla 16 diligenciada para un conjunto de prueba
3	Establecer una escala de medición y la regla de clasificación	Capítulo 9, sección 9.3	Cuatro niveles, de Inicial a Listo para IA, con regla del nivel más alto cumplido por completo (Tabla 17)	Tabla 17 y regla de decisión
4	Aplicar el índice sobre conjuntos reales del catálogo y documentar el resultado	Capítulo 6, sección 6.3	Muestra priorizada de diecisiete conjuntos con estado observado y mejora pendiente (Tabla 10)	Tabla 10 con enlaces a los conjuntos
5	Definir la ruta de mejora entre niveles	Capítulo 9, sección 9.3	Acciones concretas de tránsito entre niveles y registro del nivel como campo de la ficha	Párrafos finales de la sección 9.3
6	Diagnosticar el esquema de metadatos vigente de datos.gov.co	Capítulo 7, sección 7.1	Inventario de campos nativos y personalizados de la plataforma	Sección 7.1
7	Proponer los campos mínimos de metadatos para inteligencia artificial	Capítulo 7, sección 7.2	Quince campos con estado actual y acción propuesta, cinco de ellos por incorporar (Tabla 12)	Tabla 12
8	Definir vocabularios, taxonomías y condiciones de interoperabilidad semántica	Capítulo 7, sección 7.3	Adopción de DIVIPOLA, vocabularios controlados e identificadores persistentes	Sección 7.3
9	Prever la automatización del	Capítulo 7, sección 7.4	Articulación con la solución de rastreo y evaluación del	Sección 7.4

No.	Obligación derivada del anexo técnico	Ubicación en el E8	Producto o instrumento que la satisface	Medio de verificación
	levantamiento de metadatos		proyecto MinTIC y Universidad de Cartagena	
10	Diseñar la sección del portal orientada a soluciones y casos de uso de IA	Capítulo 7, sección 7.5	Ocho componentes funcionales con su insumo de metadatos (Tabla 13)	Tabla 13
11	Fijar el criterio de ingreso de un conjunto a la sección	Capítulo 7, sección 7.5	Nivel Preparado o superior, licencia abierta declarada y acceso por interfaz de programación	Sección 7.5
12	Documentar casos de uso de IA basados en datos abiertos	Capítulo 10, sección 10.4, y Anexo D	Cinco casos nacionales publicados (Tabla 21) y ficha normalizada de caso de uso	Tabla 21 y Anexo D
13	Identificar los insumos técnicos que requiere la implementación en el portal	Capítulo 7, sección 7.5	Relación de permisos, arquitectura, analítica y condiciones de Socrata que deben solicitarse	Último párrafo de la sección 7.5
14	Diseñar la ruta de formación para las entidades públicas que publican datos	Capítulo 8, sección 8.1	Ruta articulada con la oferta de la universidad corporativa del MinTIC (Tabla 14)	Tabla 14
15	Diseñar el programa de capacitación en datos abiertos para desarrolladores de IA	Capítulo 8, sección 8.2	Siete temas que reproducen el ciclo de vida del dato, con contenidos y justificación (Tabla 15)	Tabla 15
16	Prever certificación, comunidad de práctica y sostenibilidad del programa	Capítulo 8, sección 8.3	Certificación por competencias, comunidad entre cohortes y esquema de formación de formadores	Sección 8.3
17	Establecer el modelo de gobernanza de la estrategia	Capítulo 9, sección 9.7	Tres roles con responsabilidad y periodicidad definidas (Tabla 18)	Tabla 18
18	Definir el esquema de seguimiento con indicadores	Capítulo 9, sección 9.8, y Anexo E	Siete indicadores con fórmula, frecuencia y responsable (Tabla 19), desarrollados en fichas técnicas	Tabla 19 y Anexo E

Fuente: elaboración propia a partir del anexo técnico del proyecto IA para el Estado y del contenido del Entregable 8.

B.2. Trazabilidad con el CONPES 4144 de 2025

El mandato de política que origina el entregable reside en dos acciones del Documento CONPES 4144 de 2025. La acción 2.20 ordena publicar una estrategia de datos abiertos que fortalezca la apertura y el aprovechamiento de datos para el desarrollo de sistemas

de inteligencia artificial, y la acción 2.16 vincula la disponibilidad de datos abiertos de calidad con la viabilidad de esos sistemas (Departamento Nacional de Planeación, 2025). La Tabla B2 traduce ambas acciones a la respuesta concreta del documento y señala el producto que un evaluador de política puede exigir como evidencia.

Tabla B2. Trazabilidad de las acciones del CONPES 4144 de 2025

Acción	Contenido de la acción	Respuesta de la estrategia	Ubicación	Producto verificable
2.20	Publicar una estrategia de datos abiertos que fortalezca la apertura y el aprovechamiento de datos para el desarrollo de sistemas de inteligencia artificial	Documento completo, estructurado en tres ejes y dos guías de aplicación, una para entidades productoras y otra para reutilizadores	Capítulos 1 a 10	El Entregable 8 en su conjunto, con el objetivo general enunciado en la sección 5.2
2.20	Definir un modelo estratégico con objetivos y resultados esperados	Objetivo general, tres ejes encadenados y resultado esperado por eje	Sección 5.2, Tabla 7	Tabla 7 y su correspondencia con el Anexo E
2.20	Identificar los actores del ecosistema y su papel	Cinco categorías de actor con papel definido y capítulo que orienta a cada una	Sección 5.3, Tabla 8	Tabla 8
2.16	Vincular la disponibilidad de datos abiertos de calidad con la viabilidad de los sistemas de inteligencia artificial	Autoevaluación de aptitud para IA que convierte la calidad en una condición medible y publicable en la ficha del conjunto	Secciones 6.1 y 9.3, tablas 15 y 16	Tablas 16 y 17, y campo de nivel de aptitud en la Tabla 12
2.16	Habilitar la identificación de datos de alto valor para inteligencia artificial	Distinción entre datos de alto valor, críticos y estratégicos, con sectores priorizados y casos de uso de alto potencial	Secciones 6.2 y 6.3, tablas 9 y 10	Tablas 9 y 10
2.16	Asegurar condiciones de acceso, licenciamiento y documentación que permitan el entrenamiento de modelos	Licencia abierta declarada y acceso por interfaz de programación como requisitos de la priorización; diccionario de datos obligatorio	Secciones 7.2, 9.5 y 10.3	Tabla 12 y procedimiento del paso 5
2.16 y 2.20	Garantizar el uso responsable de los datos en sistemas de inteligencia artificial	Adopción de los principios de la Guía Ética, con cuatro conductas exigibles al reutilizador e incorporadas a los criterios de la ficha de caso	Secciones 6.5 y 10.5, y Anexo D	Anexo D, bloque de cumplimiento ético

Fuente: elaboración propia a partir del Documento CONPES 4144 de 2025 (Departamento Nacional de Planeación, 2025).

B.3. Trazabilidad con el marco normativo

La estrategia opera sobre obligaciones que ya existen para los sujetos obligados. La Tabla B3 identifica la exigencia concreta que cada instrumento normativo impone y la respuesta técnica que el documento ofrece.

Tabla B3. Trazabilidad entre las exigencias normativas y los instrumentos de la estrategia

Instrumento normativo	Exigencia concreta	Respuesta de la estrategia	Ubicación
Ley 1712 de 2014	Divulgación proactiva y publicación en formatos accesibles y reutilizables	Ruta de ocho pasos para la entidad productora, con la publicación como paso 5 y la actualización como paso 6	Capítulo 9
Resolución 1519 de 2020 y su anexo de datos abiertos	Estándares mínimos de estructura, formato, codificación y actualización	Procedimiento de publicación con formato estructurado, codificación UTF-8, títulos en la primera fila y una fila por registro	Sección 9.5
Resolución 1519 de 2020	Accesibilidad y seguridad digital de los contenidos publicados	Condición de que la sección de IA del portal respete los estándares de accesibilidad y seguridad vigentes	Sección 7.5
Decreto 767 de 2022	Gestión del dato como activo institucional y decisiones basadas en datos	Incorporación del avance de la autoevaluación como tema del comité institucional de gestión y desempeño	Sección 9.7
Plan Nacional de Infraestructura de Datos e IDEC	Lenguaje común de interoperabilidad, calidad y gobernanza	Campos mínimos de metadatos alineados con DCAT y con la guía de estándares de calidad, y vocabularios compartidos	Secciones 7.2 y 7.3
Guía Ética de IA para entidades públicas	Principios de transparencia, control humano, no discriminación y protección de datos	Cuatro conductas exigibles al reutilizador y verificación de ausencia de datos personales antes de publicar	Secciones 9.5, 10.5 y Anexo D
Ley 1581 de 2012 y régimen de reserva legal	Protección de datos personales e información clasificada o reservada	Etiqueta obligatoria de nivel de sensibilidad en la ficha del conjunto y verificación previa a la publicación	Tabla 11 y sección 9.5

Fuente: elaboración propia a partir de la Tabla 2 del documento principal y de la normativa citada.

B.4. Trazabilidad interna entre ejes, instrumentos e indicadores

La Tabla B4 recorre esa cadena para los tres ejes definidos en la sección 5.2 y remite a los códigos de indicador que el Anexo E desarrolla, de modo que la evaluación del avance de un eje se resuelve consultando un conjunto acotado de mediciones.

Tabla B4. Cadena de trazabilidad entre ejes estratégicos, instrumentos e indicadores

Eje	Capítulos	Instrumentos que lo materializan	Indicadores del Anexo E
E1. Calidad y priorización	6 y 9	Criterios mínimos de aptitud, lista de verificación de ocho aspectos, escala de cuatro niveles, muestra priorizada de diecisiete conjuntos y clasificación de datos de alto valor, críticos y estratégicos	IND-01, IND-02, IND-03
E2. Metadatos y descubribilidad	7 y 10	Quince campos mínimos de metadatos, vocabularios controlados, automatización del levantamiento, sección de IA del portal con ocho componentes y acceso programático documentado	IND-04, IND-05
E3. Capacitación y gobernanza	8 y 9	Ruta de formación para entidades, programa de siete temas para desarrolladores, certificación por competencias, comunidad de práctica, modelo de roles y tablero de seguimiento	IND-06, IND-07

Fuente: elaboración propia a partir de la Tabla 7 del documento principal y del Anexo E.

Anexo C. Machine Learning Canvas

C.1. Función y momento de aplicación

El Machine Learning Canvas es un lienzo de una página que ayuda a explicitar, antes de escribir código, las decisiones que determinan si un proyecto de aprendizaje automático tiene sentido (Dorard, 2015). Su lógica es la del modelo de negocio en una hoja, trasladada a la construcción de sistemas predictivos: cada bloque plantea una pregunta cuya respuesta condiciona a las demás, y la incoherencia entre bloques revela un proyecto mal planteado antes de que consuma recursos.

El documento principal lo sitúa en la fase de entendimiento del negocio y de los datos de la metodología CRISP-ML(Q) (Studer et al., 2021), que es la primera del ciclo descrito en la Tabla 20. Parte de los proyectos de analítica pública fracasa por una causa que se conoce tarde, cuando el equipo descubre que los datos disponibles no sostienen el objetivo prometido, que la decisión que el modelo debía apoyar no tiene doliente institucional, o que la licencia de los datos impide el uso previsto. Diligenciar el lienzo cuesta unas horas de trabajo colectivo y evita comprometer meses de desarrollo sobre supuestos que no resisten el examen.

La versión que adopta esta estrategia conserva los diez bloques de la propuesta original y añade cuatro que responden a las condiciones del sector público colombiano. La adición no es decorativa. Un sistema construido sobre datos abiertos del Estado enfrenta restricciones de licenciamiento y de protección de datos personales que un proyecto privado resuelve por contrato; hereda los sesgos de registros administrativos que reflejan desigualdades territoriales de capacidad institucional; debe explicar sus resultados ante quienes son afectados por ellos; y necesita un responsable identificado que sostenga su operación cuando termine el convenio o el contrato que lo originó.

C.2. Estructura del lienzo adaptado a datos abiertos

La Tabla C1 describe los catorce bloques del lienzo, donde la columna central plantea la pregunta que el bloque resuelve y la columna final precisa qué debe registrarse cuando la fuente de datos es el Portal Nacional de Datos Abiertos, con remisión al instrumento del documento principal que aporta la información necesaria.

Tabla C1. Bloques del Machine Learning Canvas adaptado a proyectos con datos abiertos

N.	Bloque	Pregunta que resuelve	Qué se registra cuando la fuente es datos.gov.co
1	Propuesta de valor	¿Qué problema público se resuelve, para quién y qué cambia si el sistema funciona?	Problema formulado en términos de decisión pública, población afectada y costo actual de no resolverlo
2	Decisiones que habilita	¿Qué decisión concreta toma una persona o un proceso a partir de la salida del modelo?	Decisión, responsable de tomarla, frecuencia y capacidad real de actuación, incluida su restricción de volumen
3	Tarea de predicción	¿Qué se predice, sobre qué unidad de análisis y con qué tipo de aprendizaje?	Clasificación, regresión, ordenamiento o detección de anomalías; unidad de análisis y definición precisa de la variable objetivo
4	Fuentes de datos	¿De dónde provienen los datos y en qué condiciones se obtienen?	Identificador de nueve caracteres, entidad responsable, nivel de aptitud declarado, licencia, cobertura temporal y frecuencia declarada
5	Recolección y actualización	¿Cómo se obtienen los datos de forma reproducible y con qué periodicidad se refrescan?	Consulta por la interfaz de programación de Socrata, parámetros de la consulta y verificación de la vigencia efectiva contra la fecha de actualización
6	Atributos	¿Qué variables alimentan el modelo y cómo se derivan de los datos disponibles?	Variables originales, transformaciones aplicadas, equivalencias resueltas con códigos DIVIPOLA y registro del linaje de cada atributo
7	Construcción del modelo	¿Con qué técnica, con qué frecuencia de reentrenamiento y con qué recursos se construye?	Familia de algoritmos, volumen de datos, periodicidad de reentrenamiento e infraestructura disponible en la entidad
8	Evaluación fuera de línea	¿Con qué métricas y con qué esquema de validación se decide que el modelo sirve?	Métricas alineadas con la restricción de la decisión, esquema de validación temporal y umbral mínimo aceptable
9	Generación de predicciones	¿Cuándo y cómo se producen las predicciones, y en qué formato se entregan?	Procesamiento por lotes o en línea, latencia admisible, formato de entrega y punto de integración con el proceso existente
10	Evaluación en operación y monitoreo	¿Cómo se sabe que el sistema sigue funcionando después del despliegue?	Indicadores de operación, tasa de confirmación de las salidas revisadas, vigilancia de la deriva y disparadores de reentrenamiento
11	Restricciones normativas y de datos personales	¿Qué limita jurídicamente el uso previsto de los datos?	Licencia declarada y sus obligaciones de atribución, presencia de datos personales, reserva legal y medidas de anonimización o agregación
12	Sesgos y representatividad	¿Qué poblaciones o territorios están subrepresentados y qué efecto tiene ese vacío?	Cobertura territorial e institucional del conjunto, brechas de registro conocidas y prueba de desempeño diferenciado por subgrupo

N.	Bloque	Pregunta que resuelve	Qué se registra cuando la fuente es datos.gov.co
13	Requisitos no funcionales	¿Qué exigencias de robustez, escalabilidad, explicabilidad y eficiencia debe cumplir?	Nivel de explicación exigible ante el afectado, tolerancia a fallos, volumen máximo previsto y consumo de cómputo admisible
14	Sostenibilidad y responsable	¿Quién opera el sistema, con qué presupuesto y qué ocurre cuando el equipo actual se retira?	Área responsable, costo anual estimado, documentación de traspaso y previsión de continuidad tras el contrato o convenio

Fuente: elaboración propia a partir de Dorard (2015), Studer et al. (2021) y las condiciones de reutilización descritas en el capítulo 10 del documento principal.

La coherencia entre bloques importa más que el detalle de cada uno. Un lienzo bien diligenciado permite verificar tres consistencias. La primera enlaza los bloques 2, 3 y 8, porque la métrica de evaluación debe corresponder a la restricción real de la decisión; un modelo evaluado por exactitud global resulta inútil si la entidad solo puede revisar treinta casos por semana, situación en la que la métrica pertinente mide la precisión sobre esos treinta. La segunda enlaza los bloques 4, 11 y 12, dado que la licencia y la cobertura del conjunto determinan qué usos son admisibles y sobre qué población el resultado es defendible. La tercera enlaza los bloques 5, 10 y 14, pues un sistema que consume datos por interfaz de programación puede reejecutarse cuando la serie avance, mientras que uno alimentado por descargas manuales depende de una persona que tarde o temprano cambia de cargo.

C.3. Criterios de decisión sobre la viabilidad

La Tabla C2 fija cinco criterios de corte que permiten una autoevaluación para dar viabilidad a un proyecto de machine learning. Registrar la decisión de detener un proyecto, con la razón que la sustenta, protege a la entidad y produce aprendizaje reutilizable para formulaciones posteriores.

Tabla C2. Criterios de corte para la decisión sobre la viabilidad del proyecto

Criterio	Pregunta de corte	Efecto si no se cumple
Aptitud de los datos	¿Los conjuntos alcanzan al menos el nivel Preparado de la autoevaluación, con diccionario, vigencia y acceso programático?	Se detiene el proyecto y se solicita a la entidad productora la mejora del conjunto según la ruta de la sección 9.3
Habilitación jurídica	¿La licencia declarada permite el uso previsto y el conjunto está libre de datos personales o reservados no tratados?	Se detiene el proyecto. La ausencia de licencia clara expone la iniciativa a un riesgo jurídico que no se resuelve por vía técnica
Doliente de la decisión	¿Existe un área identificada que tomará la decisión, con capacidad real de actuar sobre la salida del modelo?	Se detiene el proyecto. Un sistema sin decisión asociada produce tableros que nadie consulta

Criterio	Pregunta de corte	Efecto si no se cumple
Correspondencia entre métrica y restricción	¿La métrica de evaluación refleja la capacidad real de actuación de la entidad?	Se rediseña el bloque 8 antes de continuar, porque una métrica mal elegida valida modelos que fracasan en operación
Sostenibilidad	¿Hay responsable, presupuesto de mantenimiento y documentación suficiente para que el sistema sobreviva al equipo que lo construye?	Se rediseña el alcance hacia una versión más simple y mantenible, o se acuerda un plan de traspaso antes de iniciar el desarrollo

Fuente: elaboración propia.

Anexo D. Ficha de caso de uso de inteligencia artificial

D.1. Función de la ficha y ruta de publicación

La ficha de caso de uso cumple tres funciones dentro de la estrategia:

- Alimenta el componente de fichas de caso de la sección de inteligencia artificial del portal, descrito en la Tabla 13 del documento principal, con lo cual materializa parte del tercer componente del hito del CONPES.
- Cierra el ciclo entre reutilización y publicación, porque cada caso documentado informa a las entidades productoras qué datos generan valor y orienta su priorización.
- Constituye el producto integrador del programa de capacitación para desarrolladores descrito en la sección 8.2, de modo que la formación alimenta de manera directa el catálogo de casos en lugar de terminar en un certificado sin evidencia aplicada.

La normalización del formato responde a una oportunidad de mejora observada en la sección de usos del portal, donde los casos publicados describen los productos con niveles de detalle dispares, y buena parte omite la información que un tercero necesita para evaluar la solidez del trabajo o para replicarlo, en particular los conjuntos empleados con su identificador, las métricas obtenidas y los límites reconocidos de la solución.

D.2. Estructura de la ficha

La Tabla D1 muestra los campos recomendados para la construcción de la ficha, agrupados en ocho bloques. La columna de obligatoriedad distingue los campos exigibles para la publicación de los que enriquecen el caso sin condicionarla. La columna final indica qué componente del portal o qué metadato consume la información, lo que evita campos que nadie usa y sustenta la exigencia ante el autor.

Tabla D1. Estructura de la ficha de caso de uso de inteligencia artificial

N.º	Campo	Descripción	Destino en el portal
A. Identificación			
1	Título del caso	Nombre corto que identifique el producto y el problema que atiende	Listado y buscador de la sección de IA
2	Autor u organización	Persona, grupo de investigación, empresa o entidad responsable del desarrollo	Ficha del caso y filtro por tipo de reutilizador
3	Tipo de reutilizador	Academia, entidad pública, empresa, emprendimiento,	Filtro de la sección de usos

N.º	Campo	Descripción	Destino en el portal
		sociedad civil o periodismo de datos	
4	Sector y cobertura territorial	Sector temático y ámbito geográfico del caso, con código DIVIPOLA cuando aplique	Filtros por sector y departamento
5	Estado de desarrollo	Prototipo, piloto, en operación o descontinuado, con la fecha del estado reportado	Ficha del caso
B. Problema público			
6	Problema que atiende	Descripción del problema público en términos de decisión, población afectada y situación previa	Ficha del caso
7	Decisión o proceso que apoya	Decisión concreta que la solución informa y actor que la toma	Ficha del caso
8	Población o territorio beneficiario	Alcance estimado de la población atendida o del territorio cubierto	Ficha del caso
C. Datos empleados			
9	Conjuntos de datos abiertos utilizados	Nombre, entidad responsable e identificador de nueve caracteres de cada conjunto	Enlace bidireccional con la ficha del conjunto
10	Nivel de aptitud de los conjuntos	Nivel declarado en la autoevaluación para cada conjunto empleado	Tablero de calidad y recomendación de datasets
11	Fuentes complementarias	Datos propios, comerciales o de terceros que se integraron a los datos abiertos	Ficha del caso
12	Periodo y volumen de los datos	Cobertura temporal utilizada y número aproximado de registros	Ficha del caso
13	Preparación aplicada	Limpieza, integración, resolución de equivalencias y transformaciones realizadas	Recursos reutilizables
D. Enfoque metodológico			
14	Tipo de tarea	Clasificación, regresión, agrupamiento, ordenamiento, procesamiento de lenguaje natural, visión o sistema conversacional	Filtro por enfoque
15	Técnicas y herramientas	Familia de algoritmos, bibliotecas y entorno de ejecución utilizados	Ficha del caso
16	Esquema de validación	Partición de los datos, validación temporal o cruzada y criterio de selección del modelo	Ficha del caso
E. Resultados			

N.º	Campo	Descripción	Destino en el portal
17	Métricas de desempeño	Métricas obtenidas, con su valor y la razón por la que se eligieron	Ficha del caso
18	Comparación con la práctica previa	Desempeño frente al procedimiento que la solución reemplaza o complementa	Ficha del caso
19	Resultado o beneficio observado	Efecto documentado sobre el proceso, el tiempo, el costo o la cobertura	Historias de uso
F. Límites y riesgos			
20	Límites de la solución	Situaciones en las que el sistema no debe usarse y supuestos que condicionan su validez	Ficha del caso
21	Sesgos identificados	Poblaciones o territorios subrepresentados y medidas de mitigación aplicadas	Metadato de sesgos, aptitud y uso para IA
22	Calidad de datos observada	Errores, vacíos o inconsistencias detectados en los conjuntos empleados	Retroalimentación a la entidad productora
G. Cumplimiento ético			
23	Verificación de reidentificación	Comprobación de que los datos empleados no permiten reidentificar personas, incluso por cruce con otras fuentes	Criterio de aceptación editorial
24	Examen de sesgos previo al entrenamiento	Revisión de la representatividad de los datos antes de entrenar y decisiones adoptadas	Criterio de aceptación editorial
25	Explicabilidad ante el afectado	Forma en que los resultados pueden explicarse a quienes son afectados por ellos	Criterio de aceptación editorial
26	Control humano de la decisión	Punto del proceso en el que una persona valida o puede revertir la salida del sistema	Criterio de aceptación editorial
H. Reproducibilidad y contacto			
27	Licencias y atribución	Licencia de cada conjunto empleado, cumplimiento de la atribución y licencia del producto derivado	Ficha del caso
28	Recursos publicados	Enlace al código, al cuaderno de análisis, a la demostración o al informe técnico	Recursos reutilizables
29	Contacto y fecha de la ficha	Correo de contacto del autor y fecha de diligenciamiento o de última actualización	Ficha del caso

Fuente: elaboración propia a partir de la Tabla 13 del documento principal, de la sección 10.5 y de los principios de la Guía Ética de IA para entidades públicas (Gobierno de Colombia, 2025).

D.2. Criterios de aceptación editorial

La verificación previa a la publicación protege la utilidad de la sección. Una vitrina que acepta cualquier envío pierde su valor como material de aprendizaje y expone al portal a respaldar de manera implícita soluciones sin sustento. La Tabla D4 fija siete criterios de aceptación con su forma de verificación y la consecuencia del incumplimiento. El equipo técnico de datos aplica la revisión con la periodicidad semestral que le asigna la Tabla 18, y las devoluciones indican el criterio incumplido con una observación concreta.

Tabla D4. Criterios de aceptación editorial de la ficha de caso de uso

Criterio	Verificación	Consecuencia del incumplimiento
Trazabilidad de los datos	Cada conjunto declarado existe en el catálogo y su identificador de nueve caracteres corresponde al nombre reportado	Devolución. Sin identificador verificable no puede construirse el enlace con la ficha del conjunto
Completitud de los campos obligatorios	Los veinticuatro campos marcados como obligatorios están diligenciados con contenido sustantivo	Devolución con señalamiento de los campos vacíos o resueltos con generalidades
Declaración de métricas	El caso reporta al menos una métrica con su valor y la razón de su elección	Devolución. Un caso sin métrica no permite juzgar la solidez del trabajo
Reconocimiento de límites	El campo de límites describe situaciones concretas en las que la solución no debe usarse	Devolución. La ausencia de límites declarados suele indicar una evaluación insuficiente
Cumplimiento del bloque ético	Los cuatro campos de cumplimiento ético describen acciones realizadas y no intenciones	Devolución. El bloque traduce obligaciones de la Guía Ética y no admite respuesta formal
Cumplimiento de licencias	La atribución de cada conjunto se cumple y la licencia del producto derivado respeta las condiciones de la fuente	Devolución. Un incumplimiento de licencia publicado en el portal compromete al ministerio
Ausencia de datos personales en la salida	La solución no expone datos personales ni permite reidentificación por cruce	Rechazo. Este criterio no admite subsanación parcial

Fuente: elaboración propia.

Los casos publicados alimentan dos indicadores del Anexo E. El código IND-06 mide el número de fichas publicadas y sirve como señal de la vitalidad del ecosistema de reutilización, mientras que el campo 22 de cada ficha aporta insumos cualitativos al seguimiento de la calidad del catálogo, porque identifica defectos que la autoevaluación de la entidad productora no siempre detecta.

Anexo E. Matriz de indicadores

E.1. Estructura, convenciones y universo de medición

La Tabla 19 del documento principal define siete indicadores con su fórmula, su frecuencia y su responsable. En este sentido, este anexo convierte cada indicador en una ficha técnica calculable. El universo de medición equivale al conjunto de datos priorizados, definido como aquellos que la entidad ha seleccionado con los criterios del paso 1 y del paso 2 de la guía, es decir, los datos de alto valor, críticos y estratégicos de la sección 6.2.

E.2. Matriz consolidada de indicadores

La Tabla E1 presenta los diez indicadores con su fórmula, su unidad, su frecuencia y su responsable. Las fichas técnicas de la sección E.3 desarrollan cada una con el detalle necesario para el cálculo.

Tabla E1. Matriz consolidada de indicadores de la estrategia

Código	Indicador	Dimensión	Fórmula	Unidad y frecuencia	Responsable
IND-01	Conjuntos listos para IA	Calidad	Conjuntos en nivel Listo para IA / total de conjuntos priorizados	Porcentaje. Anual	Equipo técnico de datos
IND-02	Conjuntos preparados o superiores	Calidad	Conjuntos en nivel Preparado o Listo para IA / total de conjuntos priorizados	Porcentaje. Anual	Equipo técnico de datos
IND-03	Licencia abierta declarada	Apertura	Conjuntos con licencia abierta declarada / total de conjuntos priorizados	Porcentaje. Anual	Entidad productora
IND-04	Metadatos completos	Calidad	Conjuntos con los quince campos mínimos diligenciados / total de conjuntos priorizados	Porcentaje. Anual	Entidad productora
IND-05	Reutilización	Uso	Suma de visualizaciones y descargas registradas sobre los conjuntos priorizados en el periodo	Número. Anual	Coordinación (MinTIC)
IND-06	Casos de uso publicados	Uso	Número de fichas de caso de IA aceptadas y publicadas en la sección del portal	Número. Anual	Coordinación (MinTIC)
IND-07	Desarrolladores certificados	Capacidades	Número de participantes que culminan el programa y obtienen la certificación por competencias	Número. Anual	Coordinación (MinTIC)

Fuente: elaboración propia a partir de la Tabla 19 del documento principal.

E.3. Fichas técnicas

Cada ficha define el indicador con el nivel de precisión que exige un cálculo reproducible. Las líneas base aparecen sin valor porque la estrategia no ha ejecutado la medición inicial; establecerlas es la primera tarea del tablero de seguimiento y debe realizarse sobre el universo priorizado con corte a una fecha fija, que se recomienda coincida con el cierre de la vigencia en la que se adopte la guía. Por tanto, las metas se presentan como propuestas y quedan sujetas al valor que arroje la medición inicial.

Tabla E2. Ficha técnica del indicador IND-01. Conjuntos listos para IA

Campo	Contenido
Dimensión y eje	Calidad. Eje E1, calidad y priorización
Definición	Proporción de los conjuntos de datos priorizados que alcanzan el nivel más alto de la autoevaluación de aptitud para inteligencia artificial
Fórmula	IND-01 = (conjuntos en nivel Listo para IA / total de conjuntos priorizados) x 100
Variables y fuente	Numerador y denominador provienen del campo de nivel de aptitud para IA que la Tabla 12 propone incorporar a la ficha del conjunto, consolidado en el tablero de calidad de la sección de IA del portal
Unidad y rango	Porcentaje, de 0 a 100
Universo de medición	Conjuntos priorizados por las entidades con los criterios de los pasos 1 y 2. Se excluyen del denominador las vistas creadas por la comunidad y las capas federadas sin esquema tabular nativo
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Equipo técnico de datos
Responsable del reporte	Coordinación de la estrategia, Dirección de Gobierno Digital del MinTIC
Línea base	Por establecer en la medición inicial. La Tabla 10 del documento principal ofrece un punto de partida sobre diecisiete conjuntos de alto valor, en los que la mejora pendiente se concentra en el diccionario de datos, la vigencia y la atribución
Meta propuesta	30 por ciento al término del primer año y 70 por ciento al término del tercero, sujeto a la línea base
Desagregación	Por sector, por entidad, por orden nacional o territorial y por categoría de dato, es decir alto valor, crítico o estratégico
Interpretación y limitaciones	El indicador depende de una autoevaluación, de modo que su validez exige la verificación de una muestra por parte del equipo técnico, tal como prevé la sección 9.3. Sin esa verificación el

Campo	Contenido
	indicador tiende a sobreestimar el avance, riesgo que aumenta cuando la cifra se usa para comparar entidades

Fuente: elaboración propia.

Tabla E3. Ficha técnica del indicador IND-02. Conjuntos preparados o superiores

Campo	Contenido
Dimensión y eje	Calidad. Eje E1, calidad y priorización
Definición	Proporción de los conjuntos priorizados que alcanzan el nivel Preparado o el nivel Listo para IA, es decir, los que cumplen el umbral de ingreso a la sección de inteligencia artificial del portal
Fórmula	IND-02 = ((conjuntos en nivel Preparado + conjuntos en nivel Listo para IA) / total de conjuntos priorizados) x 100
Variables y fuente	Campo de nivel de aptitud para IA en la ficha del conjunto, consolidado en el tablero de calidad
Unidad y rango	Porcentaje, de 0 a 100
Universo de medición	Idéntico al de IND-01, con lo cual ambos indicadores son comparables entre sí y su diferencia mide la masa de conjuntos que se encuentran a un paso del nivel máximo
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Equipo técnico de datos
Responsable del reporte	Coordinación de la estrategia
Línea base	Por establecer en la medición inicial
Meta propuesta	60 por ciento al término del primer año y 90 por ciento al término del tercero
Desagregación	Por sector, por entidad y por orden administrativo
Interpretación y limitaciones	Este indicador se mueve antes que IND-01 y funciona como señal temprana del avance, porque el tránsito de En desarrollo a Preparado exige menos esfuerzo que la conquista del nivel máximo. Una brecha amplia y persistente entre IND-02 e IND-01 señala que los conjuntos se estancan en detalles finales, casi siempre el acceso programático o los códigos estándar

Fuente: elaboración propia.

Tabla E4. Ficha técnica del indicador IND-03. Licencia abierta declarada

Campo	Contenido
Dimensión y eje	Apertura. Eje E1, calidad y priorización
Definición	Proporción de los conjuntos priorizados que declaran de manera visible una licencia abierta en el campo correspondiente de la ficha
Fórmula	IND-03 = (conjuntos con licencia abierta declarada / total de conjuntos priorizados) x 100
Variables y fuente	Campo nativo de licencia de la plataforma, recuperable por la interfaz de programación del portal
Unidad y rango	Porcentaje, de 0 a 100
Universo de medición	Conjuntos priorizados. Se considera licencia abierta cualquier variante de Creative Commons que permita la reutilización con atribución
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Entidad productora, con consolidación del equipo técnico
Responsable del reporte	Coordinación de la estrategia
Línea base	Por establecer. El diagnóstico de la sección A4 del Anexo A encontró que los conjuntos revisados declaran licencia, con variantes entre Creative Commons BY y BY-SA, lo que anticipa una línea base alta
Meta propuesta	100 por ciento al término del primer año, con sostenimiento posterior
Desagregación	Por tipo de licencia declarada y por entidad
Interpretación y limitaciones	Un valor cercano al máximo desde el inicio hace que el indicador pierda poder discriminante. Conviene entonces desplazar la atención hacia la uniformidad del licenciamiento, porque la coexistencia de variantes BY y BY-SA dentro de un mismo dominio complica la combinación de conjuntos para el entrenamiento de modelos, según documenta la sección A4

Fuente: elaboración propia.

Tabla E5. Ficha técnica del indicador IND-04. Metadatos completos

Campo	Contenido
Dimensión y eje	Calidad. Eje E2, metadatos y descubribilidad
Definición	Proporción de los conjuntos priorizados que tienen diligenciados los quince campos mínimos de metadatos para inteligencia artificial definidos en la Tabla 12

Campo	Contenido
Fórmula	IND-04 = (conjuntos con los quince campos mínimos diligenciados / total de conjuntos priorizados) x 100
Variables y fuente	Esquema de metadatos del portal, con recuperación automática por la solución de rastreo y evaluación descrita en la sección 7.4
Unidad y rango	Porcentaje, de 0 a 100
Universo de medición	Conjuntos priorizados. Mientras los cinco campos nuevos no estén incorporados al esquema del portal, el indicador se calcula sobre los diez campos disponibles y se reporta esa condición junto con la cifra
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Entidad productora, con verificación del equipo técnico
Responsable del reporte	Coordinación de la estrategia
Línea base	Por establecer en la medición inicial, una vez definido el alcance de campos evaluables
Meta propuesta	50 por ciento al término del primer año y 85 por ciento al término del tercero
Desagregación	Por campo, lo que permite identificar cuáles concentran el incumplimiento, y por entidad
Interpretación y limitaciones	El indicador mide presencia y no calidad del contenido. Un campo diligenciado con texto genérico cuenta igual que uno diligenciado con precisión, limitación que se atenúa con la verificación por muestreo y con el seguimiento separado de la cobertura del diccionario mediante IND-09

Fuente: elaboración propia.

Tabla E6. Ficha técnica del indicador IND-05. Reutilización

Campo	Contenido
Dimensión y eje	Uso. Eje E2, metadatos y descubribilidad
Definición	Volumen de visualizaciones y descargas registradas sobre los conjuntos priorizados durante el periodo de medición
Fórmula	IND-05 = suma de visualizaciones del periodo + suma de descargas del periodo, sobre los conjuntos priorizados
Variables y fuente	Métricas de uso de la plataforma, disponibles en la metadata técnica de cada conjunto y en los reportes del portal (MinTIC, 2026a)

Campo	Contenido
Unidad y rango	Número absoluto por trimestre, con cálculo adicional de la variación porcentual respecto del mismo trimestre del año anterior
Universo de medición	Conjuntos priorizados. Las cifras acumuladas que reporta la plataforma deben convertirse en flujos del periodo mediante diferencia entre cortes consecutivos
Frecuencia y corte	Anual
Responsable del cálculo	Coordinación (MinTIC)
Responsable del reporte	Coordinación de la estrategia
Línea base	Por establecer. La sección A7 del Anexo A documenta el orden de magnitud del extremo superior, con 1.736.841 visualizaciones y 208.883 descargas acumuladas en SECOP II, Contratos Electrónicos
Meta propuesta	Crecimiento del 15 por ciento anual sobre la línea base, con énfasis en los sectores de menor uso
Desagregación	Por sector, por entidad y por conjunto. Se recomienda calcular además un índice de concentración que mida qué proporción del uso total aportan los diez conjuntos más consultados
Interpretación y limitaciones	El agregado esconde la concentración documentada en el diagnóstico, en la que un grupo reducido de conjuntos absorbe la mayor parte del uso. Un crecimiento del total impulsado por conjuntos que ya eran los más consultados no refleja la ampliación de la reutilización que la estrategia persigue, razón por la cual el índice de concentración debe reportarse junto al valor absoluto

Fuente: elaboración propia.

Tabla E7. Ficha técnica del indicador IND-06. Casos de uso publicados

Campo	Contenido
Dimensión y eje	Uso. Eje E3, capacitación y gobernanza
Definición	Número de fichas de caso de uso de inteligencia artificial aceptadas y publicadas en la sección del portal durante el periodo
Fórmula	IND-06 = número de fichas publicadas en el periodo, con reporte adicional del acumulado
Variables y fuente	Registro editorial de la sección de inteligencia artificial del portal, alimentado por el procedimiento del Anexo D

Campo	Contenido
Unidad y rango	Número absoluto por año
Universo de medición	Fichas que superaron los siete criterios de aceptación de la Tabla D4. Las devueltas no se contabilizan hasta su publicación efectiva
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Coordinación (MinTIC)
Responsable del reporte	Coordinación de la estrategia
Línea base	Por establecer. La Tabla 21 del documento principal identifica cinco casos nacionales con componente de inteligencia artificial publicados en la sección de usos, cifra que puede tomarse como referencia previa a la adopción de la ficha
Meta propuesta	Diez casos publicados en el primer año y cuarenta acumulados al término del tercero
Desagregación	Por tipo de reutilizador, por sector, por tipo de tarea de inteligencia artificial y por origen, distinguiendo los casos que provienen del programa de formación
Interpretación y limitaciones	El indicador mide actividad documentada y no impacto. Un aumento del número de fichas con criterios de aceptación laxos deteriora la utilidad de la sección, por lo cual conviene reportar junto al número la proporción de fichas devueltas, que informa sobre el rigor de la revisión

Fuente: elaboración propia.

Tabla E8. Ficha técnica del indicador IND-07. Desarrolladores certificados

Campo	Contenido
Dimensión y eje	Capacidades. Eje E3, capacitación y gobernanza
Definición	Número de participantes que culminan el programa de capacitación para desarrolladores y obtienen la certificación por competencias
Fórmula	IND-07 = número de participantes certificados en el periodo, con reporte del acumulado y de la tasa de culminación
Variables y fuente	Registros de la universidad corporativa del MinTIC y del programa descrito en la sección 8.2
Unidad y rango	Número absoluto por año; la tasa de culminación se expresa en porcentaje sobre los inscritos

Campo	Contenido
Universo de medición	Participantes que completan el proyecto integrador, requisito de la certificación según la sección 8.3
Frecuencia y corte	Anual, con corte al 31 de diciembre
Responsable del cálculo	Coordinación (MinTIC)
Responsable del reporte	Coordinación de la estrategia
Línea base	Cero al inicio, dado que el programa con certificación por competencias no ha operado
Meta propuesta	Ciento cincuenta certificados en el primer año y seiscientos acumulados al término del tercero, con una tasa de culminación superior al 60 por ciento
Desagregación	Por región, por género, por tipo de vinculación laboral y por cohorte
Interpretación y limitaciones	El número de certificados mide el egreso del programa y no su efecto. La relación con IND-06 aporta esa lectura, porque el proyecto integrador culmina en una ficha de caso susceptible de publicación; una cifra alta de certificados con pocas fichas publicadas indica que la formación no está alcanzando su propósito aplicado

Fuente: elaboración propia.