



Departamento Nacional de Planeación

www.dnp.gov.co

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS**
BIG DATA
DOCUMENTO CONPES 3920

Iván Antonio Mantilla Gaviria
Director de Desarrollo Digital

 @IAMantillaG

Junio, 2018
dnp.gov.co



AGENDA - Junio, 2018

1. Antecedentes
2. Diagnóstico y definición de la política
3. Metas de la política
4. Avances y resultados



— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS**
BIG DATA
DOCUMENTO CONPES 3920

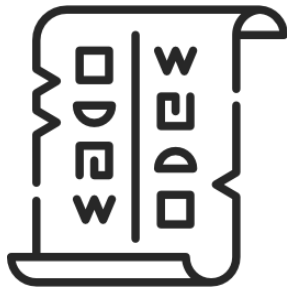
Antecedentes

1

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS
BIG DATA**
DOCUMENTO CONPES 3920

De las observaciones manuales a la primera digitalización estructurada

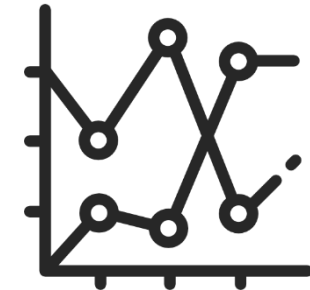
La recolección manual de datos para el análisis de fenómenos era compleja y costosa. Las primeras digitalizaciones y bases de datos estructuradas surgieron en los ochenta revelando, de manera parcial, algunos patrones de la realidad.



Procesos manuales de registro de eventos, altamente costosos: La biblioteca de Alejandría tenía 490.000 libros equivalentes a 64 Gigabytes



Recolección y digitalización selectiva y estructurada



Capacidad de análisis limitada a los datos estructurados mínimos requeridos

Fuente: Elaboración propia

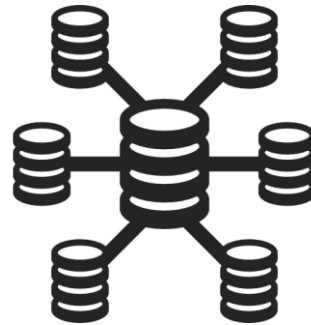
Masificación de la digitalización de los eventos

La disponibilidad de capacidades para aprovechar los datos masivos digitales que son generados por todas las interacciones humanas permite generar valor social y económico

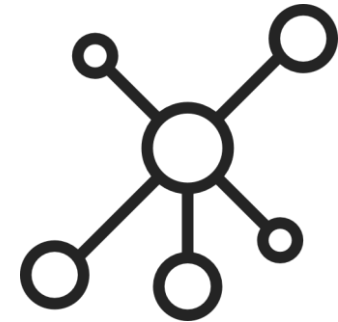
01100
10110
11110

Disponibilidad de datos masivos en formato digital a bajo costo.

Se estima que en 2020, en promedio, cada persona en el mundo generará 1,7 Megabits cada segundo



Capacidades para almacenar y explotar datos no estructurados y estructurados



Identificación de patrones y generación de conocimiento que antes era imposible por su alta complejidad. Generación de valor social y económico

Fuente: Forbes

“Datificación”: transformación del mundo en datos procesables y cuantificables

La “datificación” representa una oportunidad para generar mayor información y conocimiento, que se traduce en mayor valor social y económico para el país.

2017



1,9 billones
búsquedas



84,4 billones
correos
enviados



241 mil
millones
tweets



2,2 billones
videos vistos



25 mil millones
fotos subidas

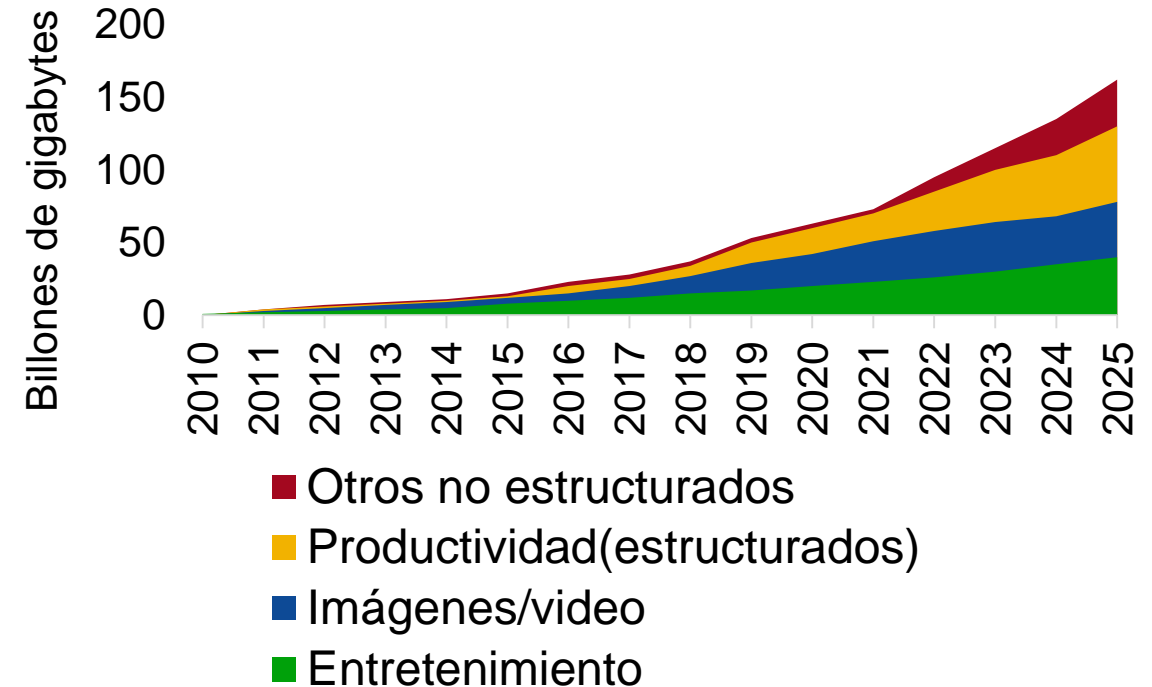


32,2 millones
páginas
hackeadas



89 mil millones
llamadas

Crecimiento de datos en el mundo

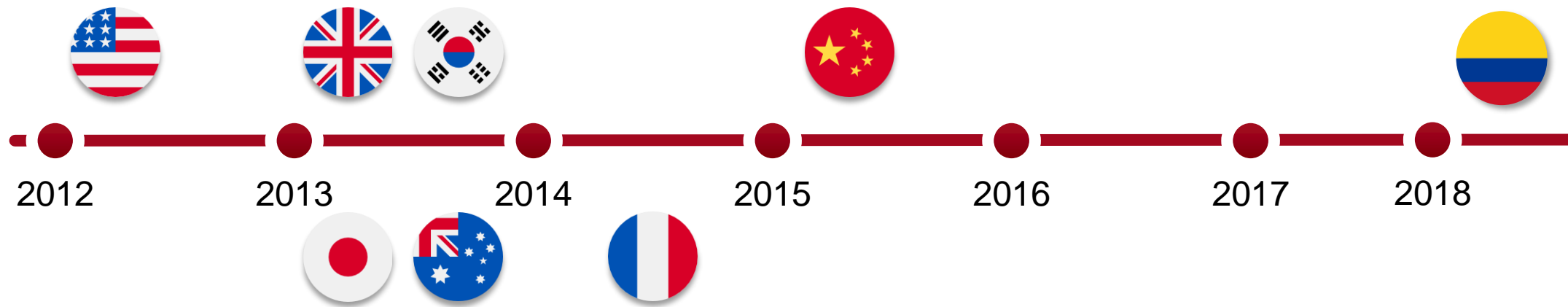


Fuente: IBM, *OCDE 2015. Internet Live Stats

Política de explotación de datos: liderazgo regional

Colombia es el primer país en América Latina y el octavo en el mundo en adoptar una política nacional de explotación de datos

Las políticas de explotación de datos buscan proveer condiciones habilitantes para lograr la generación de valor social y económico

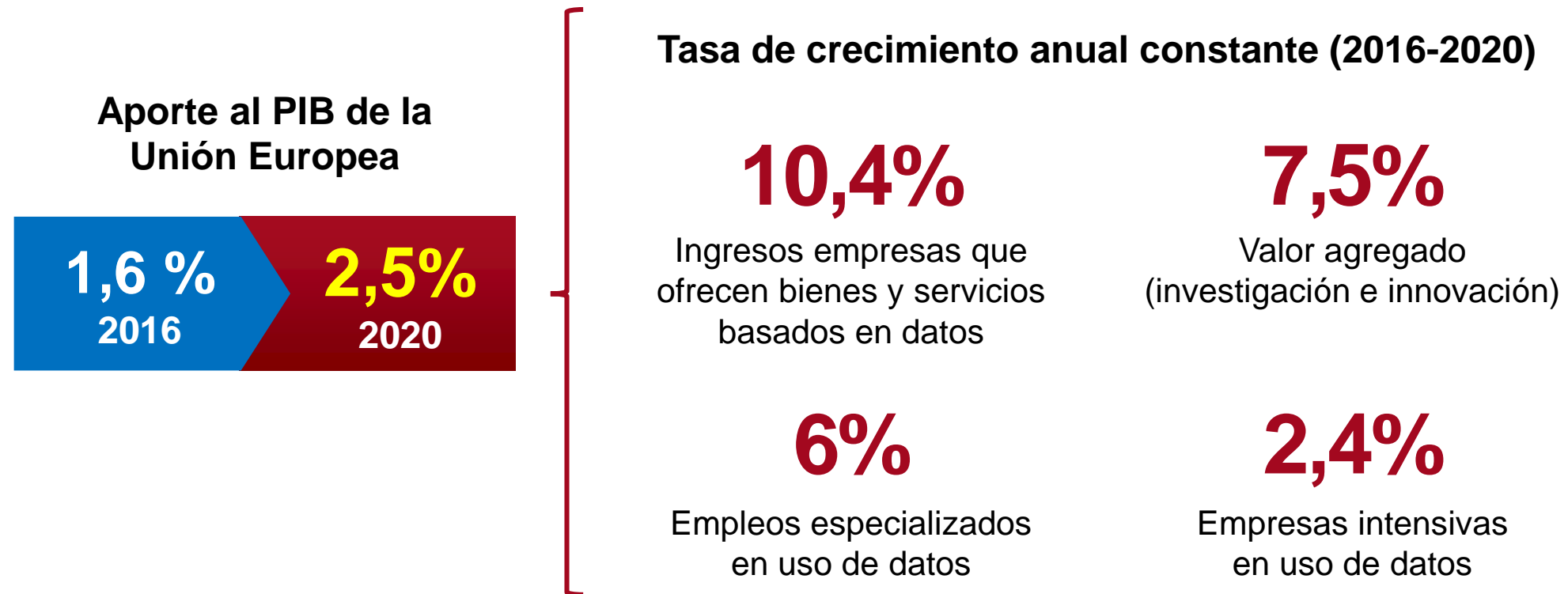


La Comisión Europea emitió una comunicación al Parlamento Europeo para que sea adoptada por sus países miembros (COM 2014 442).

* En otros países se hace referencia al uso de datos como parte de otras políticas, pero no son una política autónoma e integral.

Explotación de datos con *Big Data* tiene impactos económicos positivos

En 2020, *Big Data* representará el 2,5% del PIB de la Unión Europea.



Fuente: Comisión Europea (2017). *2020 estimado

Antecedentes y necesidad de la política de explotación de datos

El PND 2014-2018 le asignó al DNP la responsabilidad de diseñar la política de explotación de datos (*Big Data*) para el Estado. Esta es una herramienta que aporta al cumplimiento de los ODS



Desarrollo de la innovación en los países con un entorno normativo propicio a la adición de valor a los productos básicos (9.b)



Promoción del acceso público a la información (16.10)



Aumento significativo de datos oportunos, fiables y de alta calidad (17.18)

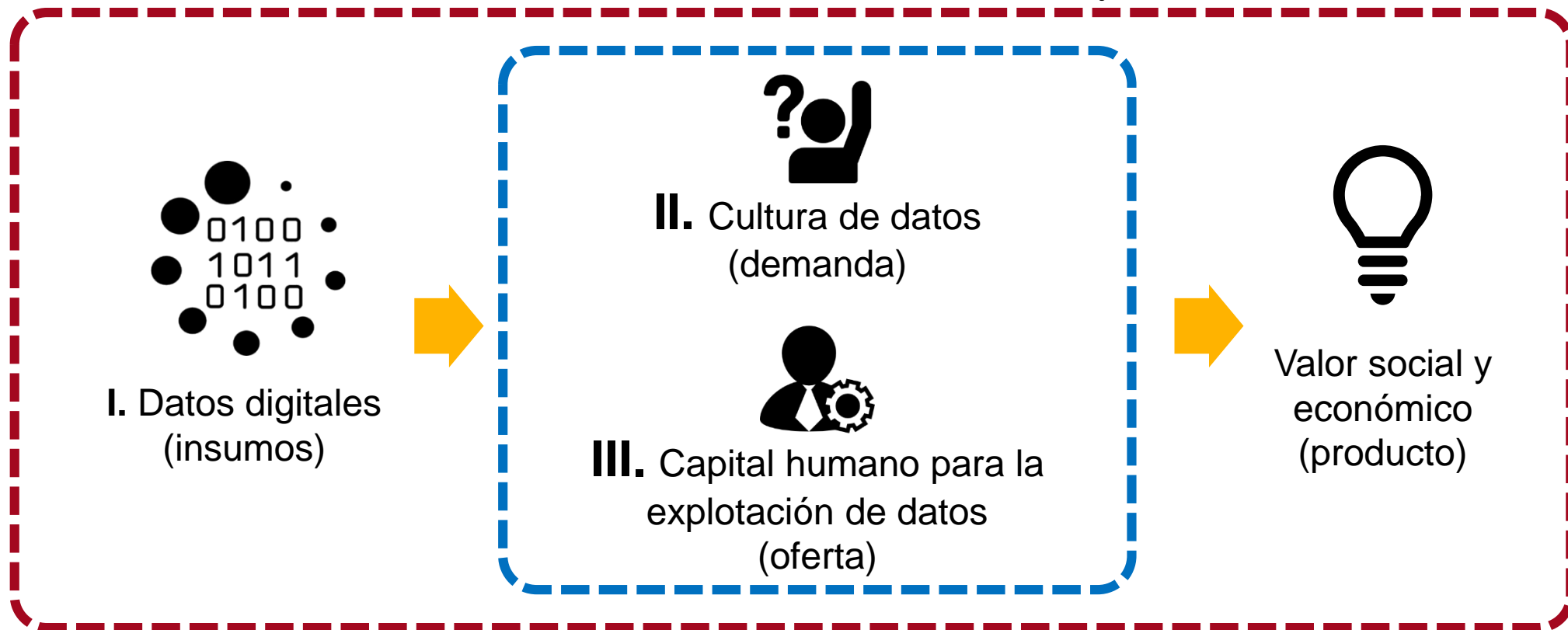
Colombia tiene retos para la medición del 46 % de los indicadores globales de ODS que pueden superarse con *Big Data*



La explotación de datos es el aprovechamiento de los mismos para generar valor social y económico

La política contenida en el Documento CONPES 3920, aborda cuatro ejes estructurales

IV. Marco jurídico, ético e institucional



Diagnóstico y definición de la política

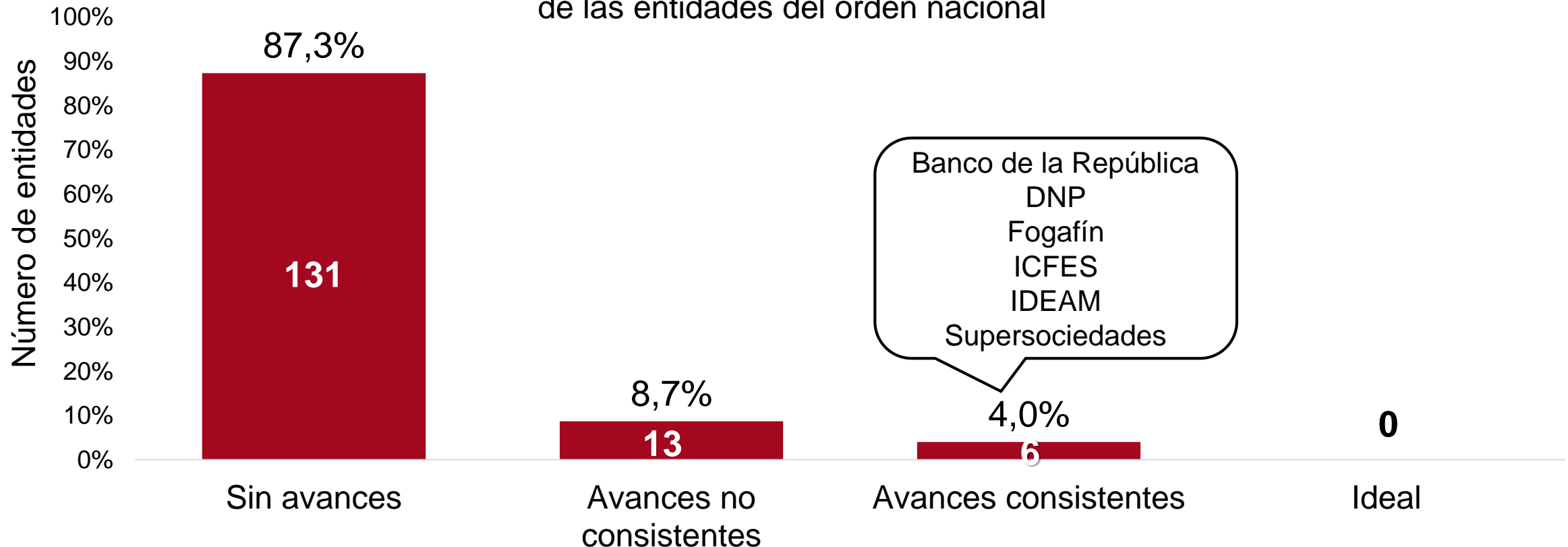
2

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS
BIG DATA**
DOCUMENTO CONPES 3920

Niveles de preparación para la explotación de datos

El 87,3 % de las entidades se encuentra sin avances en las condiciones mínimas de preparación para la explotación de datos

Niveles de preparación para la explotación de datos de las entidades del orden nacional



Fuente: Encuesta DNP (2017), Elaboración propia. Universo: 209 entidades (orden nacional). Muestra: 150 entidades.

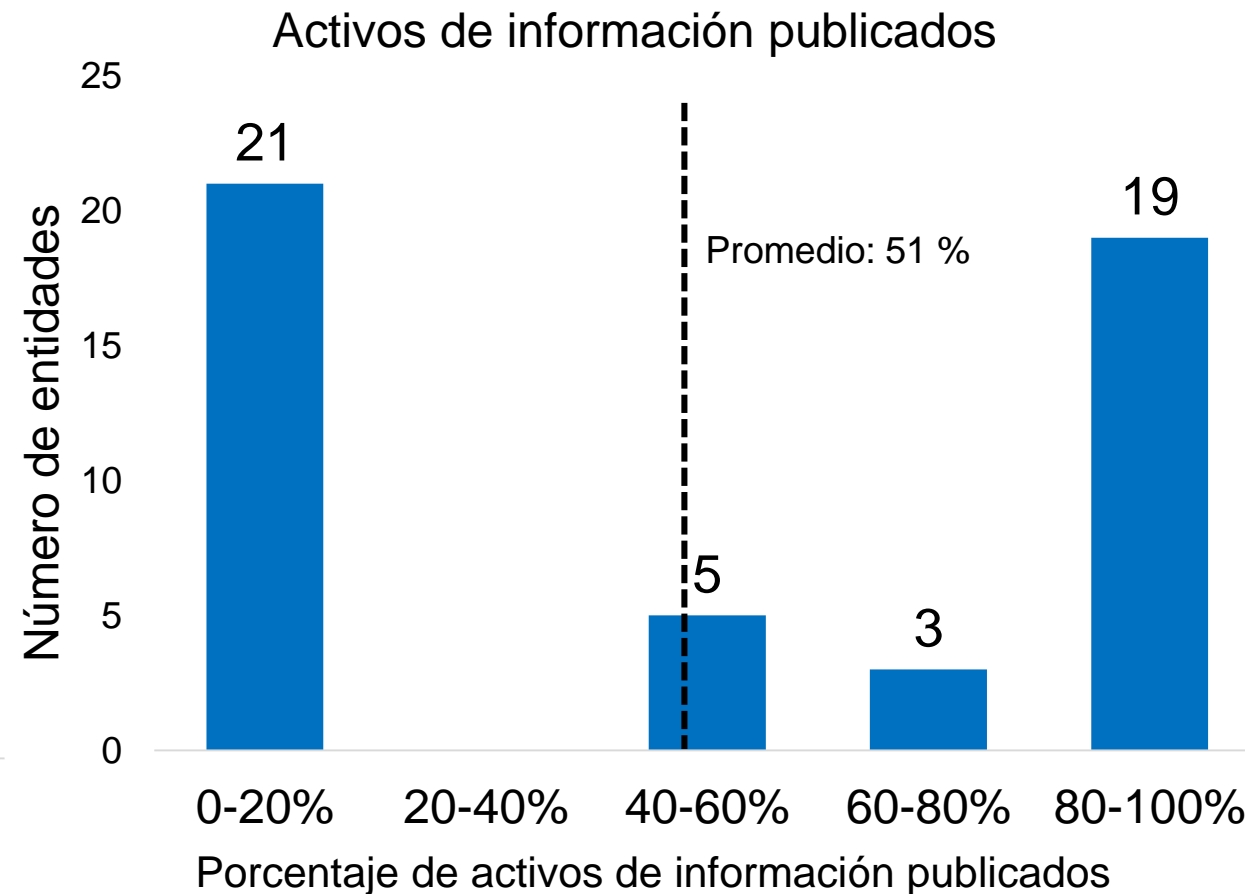
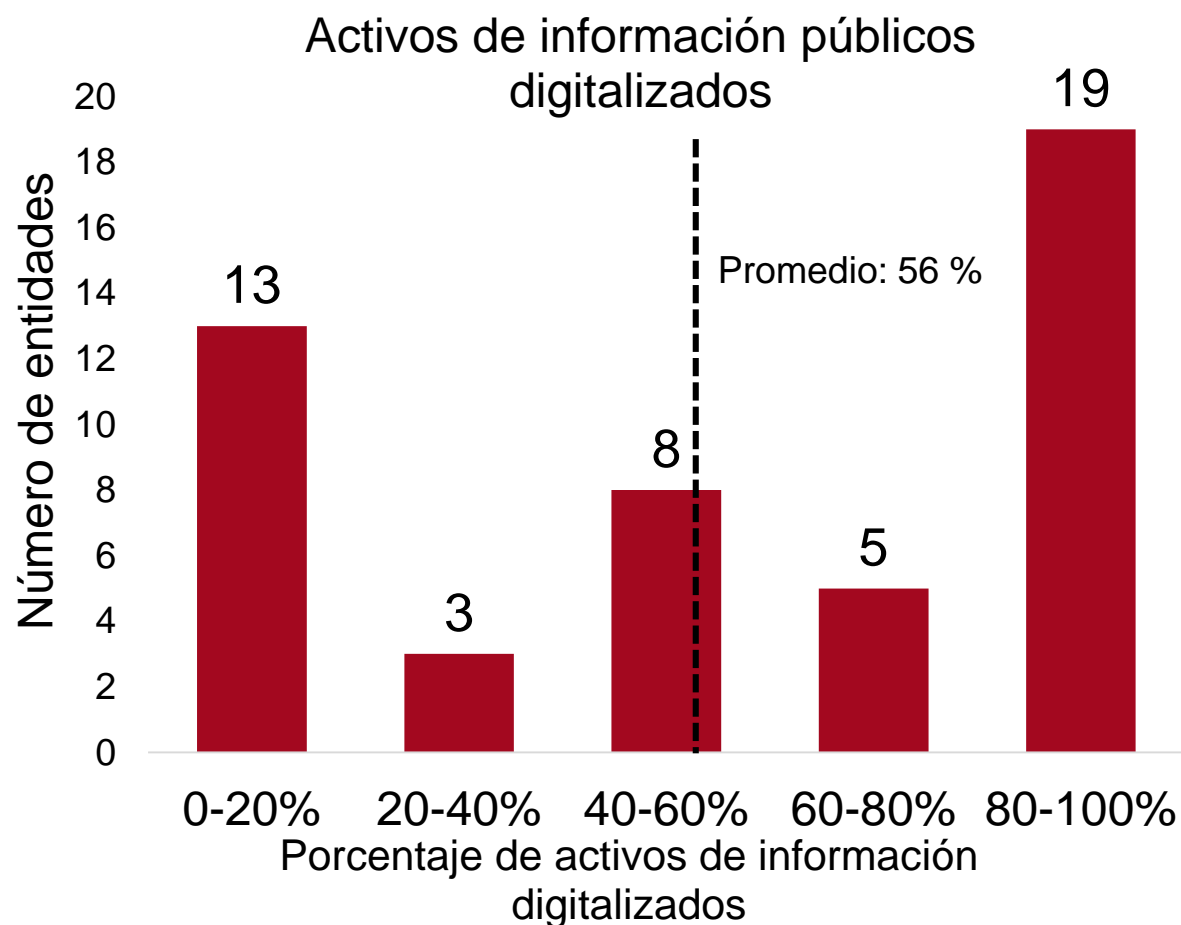
I. Insumos: Datos digitales

IV. Marco jurídico, ético e institucional



I. Insumos: Baja digitalización y apertura de datos

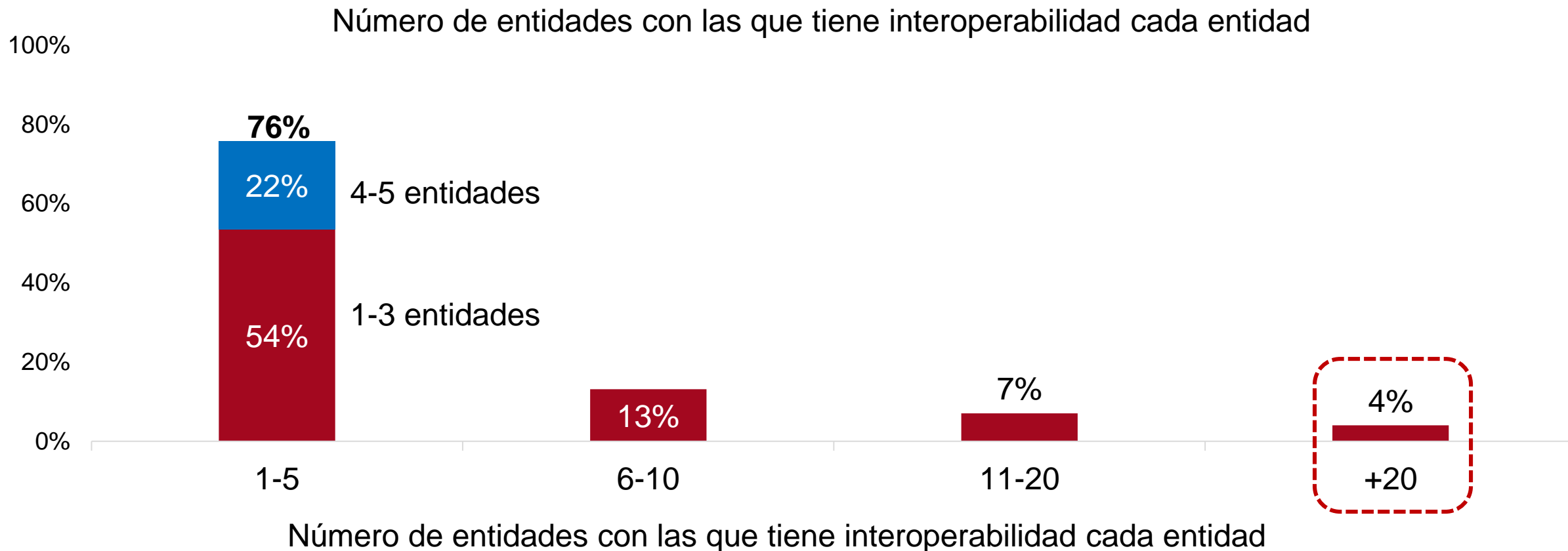
En promedio el 56 % de los activos de información pública son digitales y el 51 % de estos está publicado en la web



Fuente: Encuesta DNP (2017). *Información de 48 entidades.

I. Insumos: Baja interoperabilidad

El 66 % de entidades ha implementado algún esquema de interoperabilidad. De estas, el 4 % tiene interoperabilidad con 20 o más entidades

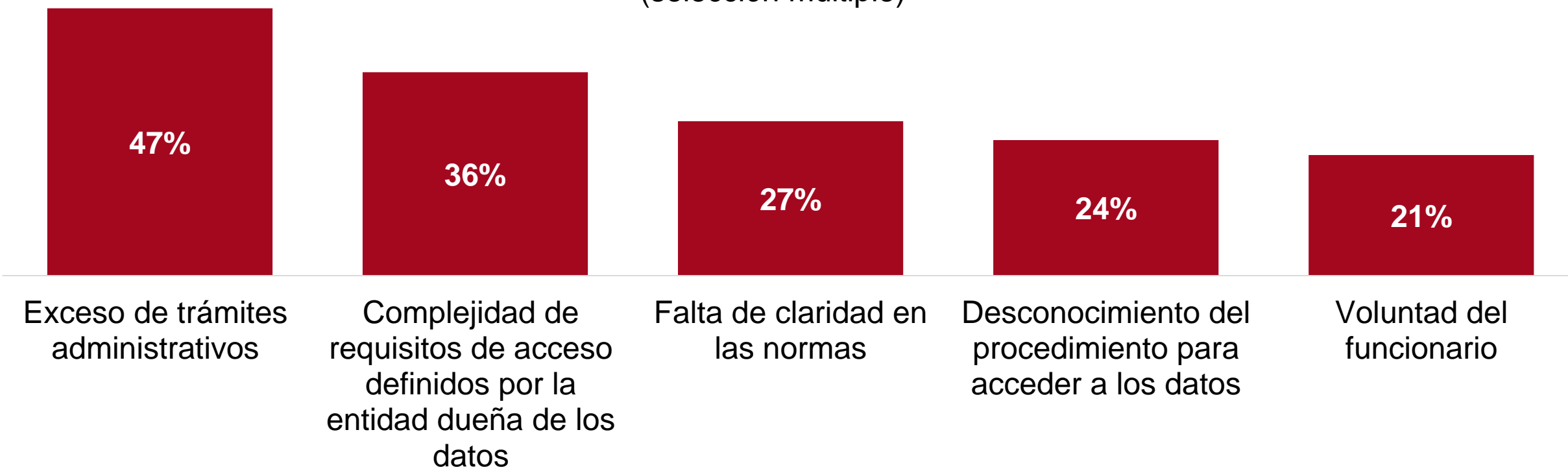


Fuente: Encuesta DNP (2017), Elaboración propia. Universo: 209 entidades (orden nacional). Muestra: 150 entidades.

I. Insumos: Ausencia de mecanismos de implementación

La ausencia de herramientas para implementar los mandatos de digitalización, apertura e interoperabilidad, se evidencia en el exceso de trámites y la complejidad de requisitos para acceder a datos públicos

Barreras para disponer e intercambiar datos en el sector público
(selección múltiple)



Fuente: Encuesta DNP (2017), Elaboración propia. Universo: 209 entidades (orden nacional). Muestra: 150 entidades.

Solución: Digitalización de los procedimientos y gestión de las entidades

Diseño e implementación de estándares y remoción de barreras que hagan efectivos los mandatos legales previamente establecidos

Estándares y herramientas

● **Desarrollar 8 estándares técnicos:** captura, anonimización, calidad, publicación, interoperabilidad, descarga, mecanismo de consumo, integración. 2018. (MinTIC, Archivo General de la Nación)
● **Crear un sistema de información para la administración de los datos de las entidades públicas** (Ley 1712 de 2014), que incluya un indicador público para hacer seguimiento a la publicación. 2019. (MinTIC)

Plan de digitalización

● **Diseñar de un plan de digitalización y automatización** de procedimientos y procesos administrativos. 2018. (Archivo General de la Nación).

Instrumento jurídico

● **Proponer Decreto reglamentario para el intercambio de datos** (artículo 159 Ley 1753 de 2015): remoción de barreras operativas e implementación de los estándares y procedimientos. 2018. (DNP, MinTIC)

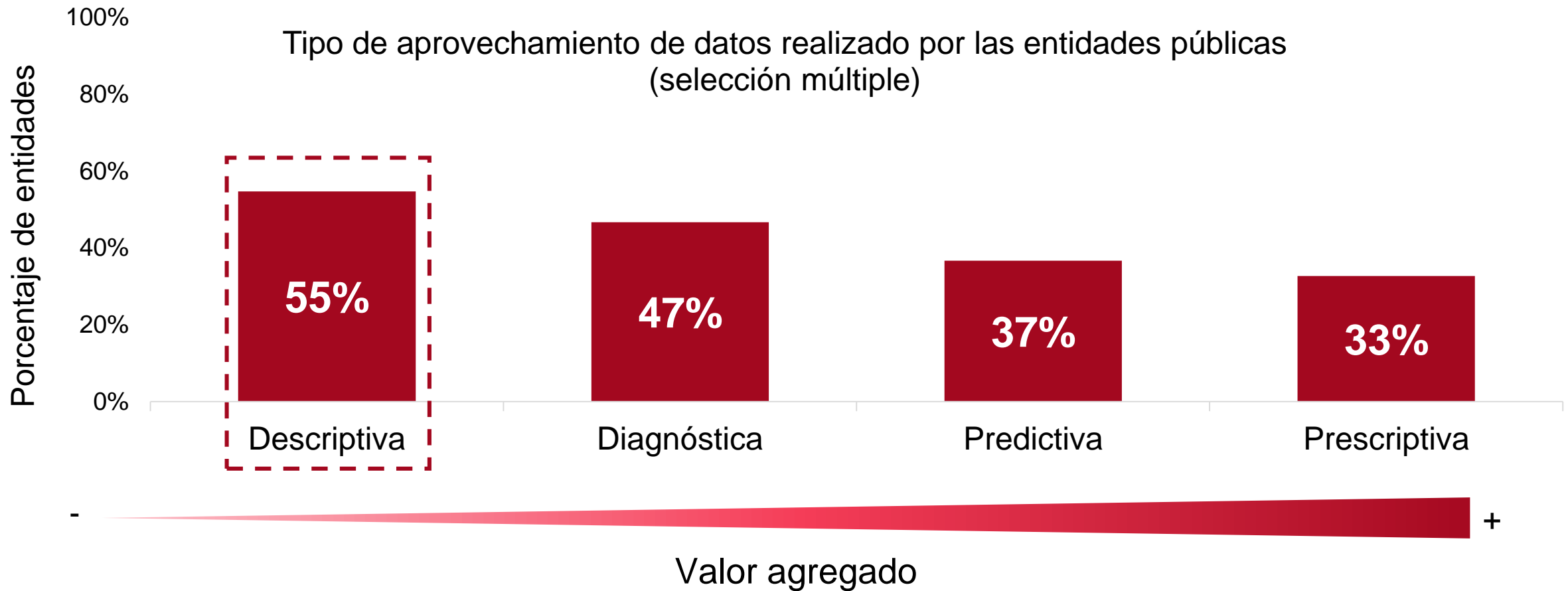
II. Cultura de datos

IV. Marco jurídico, ético e institucional



II. Cultura de datos: Explotación de datos incipiente

El aprovechamiento de datos en el Estado colombiano se focaliza principalmente en las de menor valor agregado, realizada por el 55 % de las entidades



Fuente: Encuesta DNP (2017), Elaboración propia. Universo: 209 entidades (orden nacional). Muestra: 150 entidades.

Solución: Planeación y gestión pública basada en explotación de datos

Implementación de instrumentos generadores de demanda de bienes y servicios basados en datos en el sector público

Medición y nivelación de preparación

Definición de una metodología para medir la preparación de las entidades e implementación de un plan de preparación para la explotación de datos. 2018-2019. (DNP, MinTIC).

Definición de una metodología para medir el retorno de la inversión pública de la explotación de datos. 2018. (DNP).

Contratación

Creación de un acuerdo marco de precios que facilite la adquisición de bienes y servicios para la explotación de datos. 2018. (Colombia Compra Eficiente).

Inclusión en procesos

Incorporación de la explotación de datos en la formulación de la inversión pública, la lucha contra la corrupción y el ciclo de defensa jurídica. 2018. (DNP, Secretaría de Transparencia, ANDJE).

III. Bajo capital humano para la explotación de datos

IV. Marco jurídico, ético e institucional

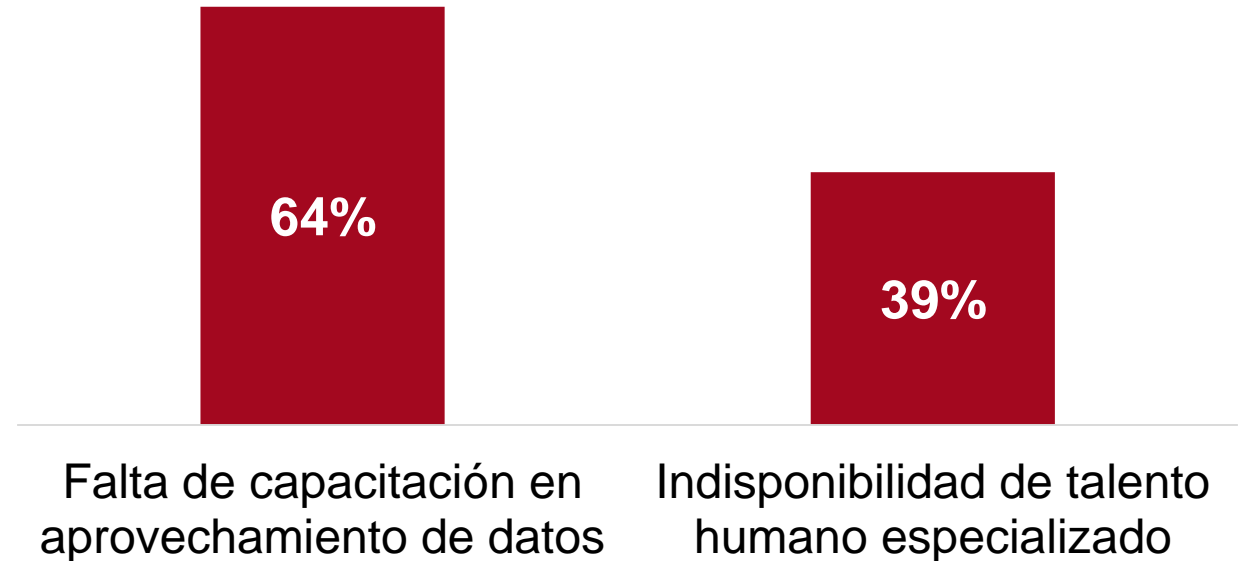
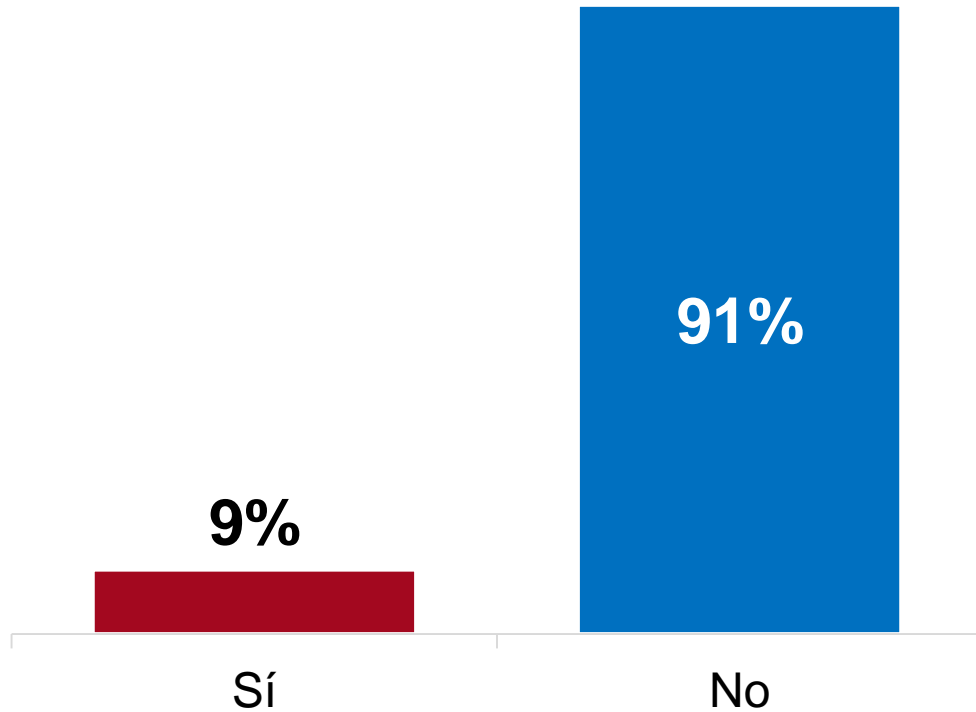


III. Capacidades: Bajo capital humano

La falta de capacitación y de disponibilidad de talento humano limitan la explotación de datos

Entidades con al menos un proyecto de explotación de datos

Barreras para la implementación de explotación de datos en las entidades (respuesta múltiple)



Fuente: Encuesta DNP (2017). "sí" incluye únicamente a las entidades respecto de las que se validó el uso de algoritmos.

Solución: Fortalecer las capacidades del capital humano en entidades públicas

Actualización transversal de competencias

Medición y cierre de brecha

- Medición de la brecha en explotación de datos. 2018. (MinTIC)
- Inclusión de competencias relacionadas con explotación de datos en el marco de cualificaciones. 2019. (MEN)

Entrenamientos sector público

- Implementación de entrenamientos en explotación de datos para 1.200 colaboradores públicos. 2018. (MinTIC)
- Inclusión de la explotación de datos en el componente de capacitación de las entidades públicas del orden nacional. 2018. (DAFP)

Emprendimiento

- Definición e implementación de una estrategia de emprendimiento en explotación de datos que permita su consolidación en el aparato productivo. 2018. (MinCIT a través de iNNpulsa -Fiducoldex)

IV. Marco jurídico, ético e institucional

IV. Marco jurídico, ético e institucional



IV. Marco jurídico, ético e institucional: Incertidumbre para la explotación

Carencia de un principio unificador y surgimiento exponencial de nuevas situaciones, generan dispersión y desactualización del marco jurídico. Esto implica retos a los límites éticos

Desarticulación



Fuente: DNP (2017)

IV. Marco jurídico, ético e institucional: Incertidumbre para la explotación

Carencia de un principio unificador y surgimiento exponencial de nuevas situaciones, generan dispersión y desactualización del marco jurídico. Esto implica retos a los límites éticos

Nuevos riesgos



Privacidad e
intimidad



Protección del
ciudadano



Transacciones y
propiedad de los datos






Monopolios

Ausencia de prácticas comunes de autorregulación

Fuente: DNP (2017)

IV. Marco jurídico, ético e institucional: Dispersión institucional

Competencias asignadas a varias entidades dentro de los regímenes relacionados con la explotación de datos, dan lugar a fallas de coordinación

	Transparencia y acceso	Eficiencia administrativa	Gobierno electrónico	Gestión documental	Protección derechos
					●
	●				
	●				●
	●	●	●		
	●	●	●		
	●	●	●	●	
	●	●	●	●	

Solución: Marco jurídico e institucional ajustado al contexto actual

Proponer las condiciones jurídicas e institucionales

Herramientas jurídicas

Elaboración de una guía para garantizar la protección de datos personales en las entidades públicas y actualización de lineamientos de contratación para incorporar la explotación de datos (guía de competencia, guía de elaboración de estudios del sector y Circular única). 2018. (SIC, Colombia Compra Eficiente, MinTIC).

Modelo de gestión

Articulación del modelo de gestión para la explotación de datos con el modelo integrado de planeación y gestión. 2019. (DAFP, DNP).

Cambio estructural

Formulación de una propuesta de norma para ajuste al marco jurídico e institucional. 2018. (DNP, MinTIC, DAFP)

Metas de la política

3

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS
BIG DATA**
DOCUMENTO CONPES 3920

Metas de la política

3

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS
BIG DATA**
DOCUMENTO CONPES 3920

Indicadores de resultado

En 2022 el 90 % de las entidades públicas del orden nacional tendrá al menos un proyecto de aprovechamiento de datos

Indicadores de resultado	2017	2020	2022
Promedio de activos públicos digitalizados y publicados	51 %	80 %	100 %
Porcentaje de entidades que tienen al menos un proyecto de aprovechamiento de datos	9,3 %	45 %	90 %
Entidades del orden nacional en el tercer nivel (sobre 4) de preparación	8,7 %	20 %	50 %

La política tiene un costo total aproximado de
16.728 millones de pesos

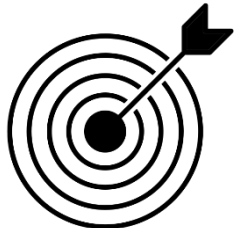
Avances y resultados

4

— POLÍTICA DE —
**EXPLOTACIÓN
DE DATOS**
BIG DATA
DOCUMENTO CONPES 3920

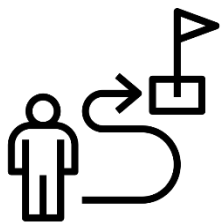
1. DNP+: Predicción del PIB Colombiano con Big Data

Herramienta del DNP basada en *Google Trends* para predecir la tendencia del PIB con tres meses de anticipación a la estadística oficial



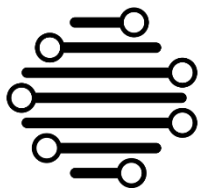
Objetivo

Conocer mensualmente, de manera anticipada, el comportamiento del PIB para tomar mejores decisiones económicas



Metodología

Metodología de pronóstico de corto plazo de series económicas para construir los indicadores que son determinados con los resultados de búsqueda que mejoran el ajuste

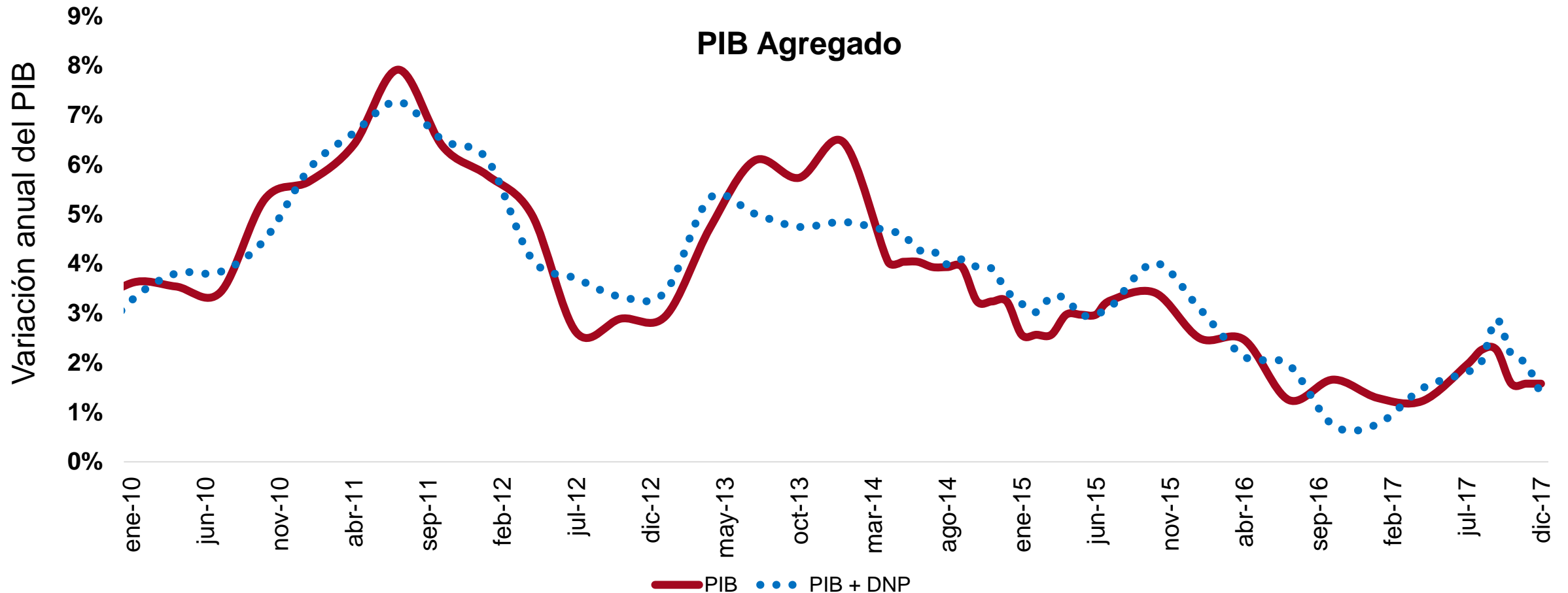


Insumos

Google Trends Estadísticas de búsquedas en el país de acuerdo con términos centrales de cada sector económico que son medidos en las cuentas nacionales

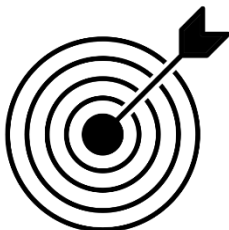
Variables exógenas representativas de la oferta en 10 sectores del PIB

1. DNP+: Predicción del PIB Colombiano con Big Data



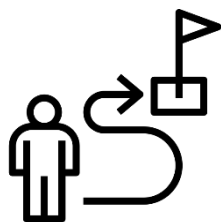
2. Big Data para caracterización de sectores en PND (1990-2014)

Medición automatizada de las prioridades temáticas en los planes nacionales de desarrollo



Objetivo

Categorizar automáticamente el texto de los Planes Nacionales de Desarrollo para facilitar el seguimiento de las prioridades sectoriales y de política pública



Metodología

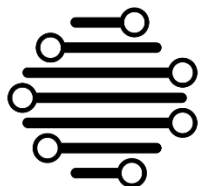
Algoritmos de procesamiento natural de lenguaje

- Lectura automática de 8.470 páginas

Análisis proyectivo de patrones en el espacio del texto

- Generación de catálogos de palabras por sector

Categorización de los planes de desarrollo y ordenación de términos por sector



Insumos

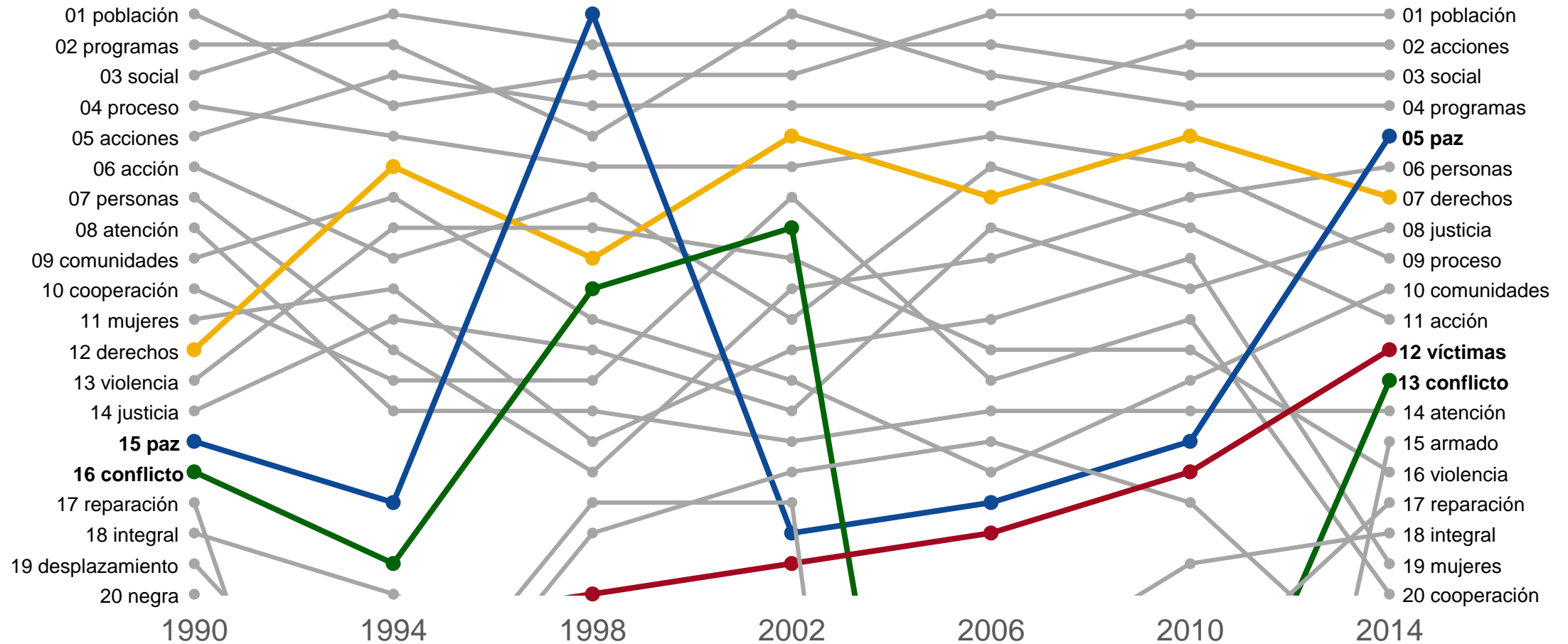
- 132 Documentos CONPES: 1.457.389 palabras en 5.025 páginas
- 7 Planes Nacionales de Desarrollo: 1.345.834 palabras en 3.445 páginas

Total: 2.803.223 palabras que caracterizan 12 sectores

2. Caracterización del sector Inclusión y Reconciliación

Términos como 'paz', 'conflicto', 'derechos' y 'víctimas' evidencian la evolución del énfasis y el aumento de las acciones desarrolladas en el marco de la finalización del conflicto.

Participación de términos en los PND para Inclusión y Reconciliación



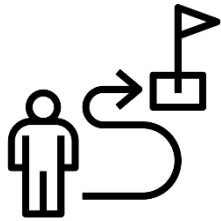
3. Big Data para caracterizar los Planes de Desarrollo Territorial

Automatización de la evaluación de la alineación de los planes de desarrollo territorial (PDT) con el plan nacional de desarrollo



Objetivo

Automatizar el seguimiento y evaluación de los planes de desarrollo territorial para optimizar el proceso de acompañamiento realizado por el DNP



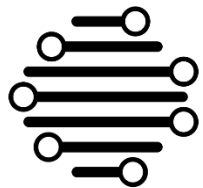
Metodología

Algoritmo de procesamiento natural del lenguaje

- Vectorización de los textos en función de la frecuencia de ocurrencia de las palabras claves de alineación

Construcción de un índice de alineación

- Comparación geométrica de la similitud de los vectores de los PDT



Insumos

146 Planes de Desarrollo Territorial

29.500 páginas

236.706 palabras totales

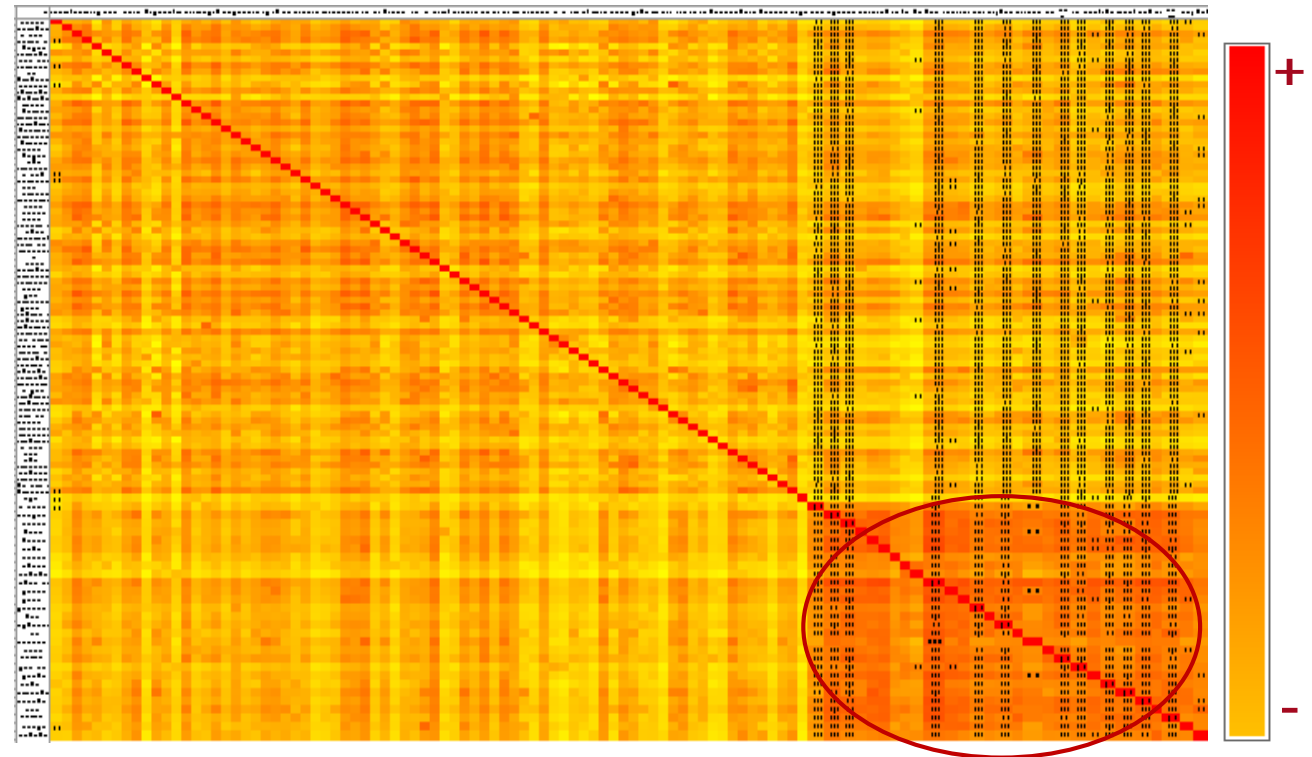
16.257 palabras únicas

3. Big Data para caracterizar los Planes de Desarrollo Territorial

Se automatiza el proceso que evalúa la alineación de los planes de desarrollo territorial con el Plan Nacional de Desarrollo. Esto generaría ahorros equivalentes a 3,8 años de trabajo humano

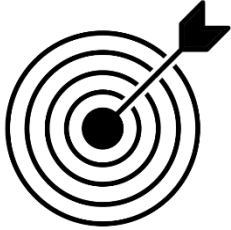
Planes de Desarrollo según alineación con Plan Nacional de Desarrollo		
Puesto	Plan de Desarrollo	Similitud
1	Cundinamarca	74%
2	Manizales (Caldas)	69%
3	Guajira	68%
4	Caldas	68%
5	Antioquia	66%
100	Carmen del Darién (Chocó)	34%
101	Santa Barbara Iscuandé (Nariño)	33%
102	Puerto Guzmán (Putumayo)	31%
103	Puerto Leguízamo (Putumayo)	31%
104	Córdoba (Bolívar)	30%

Índice de similitud



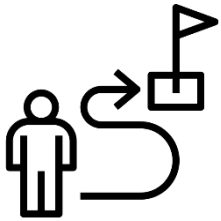
4. Big Data para el monitoreo de precios agropecuarios

Aplicación de *web scraping* diario que permite automatizar el análisis de los precios en 22 centrales mayoristas del país



Objetivo

Identificar comportamientos anómalos que puedan impactar a los pequeños productores



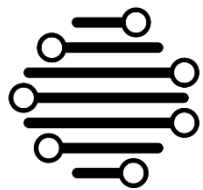
Metodología

Algoritmo que aplica *web scraping* para

- Automatizar la consulta del Sistema de Información de Precios (SIPSA)
- Hacer seguimiento en tiempo real a los precios de los productos agrícolas

Algoritmos de predicción en series de tiempo:

1. *Generalized Additive Model* (GAM),
2. TBATS,
3. *Naive*,
4. *Seasonal Naive*,
5. *Autoregressive Integrated Moving Average* (ARIMA),
6. *Error Trend Seasonality* (ETS),
7. *Mean Forecast*



Insumos

36 productos agrícolas

22 centrales mayoristas

1.296 archivos de 2013 a 2018 indexados automáticamente en una base estructurada

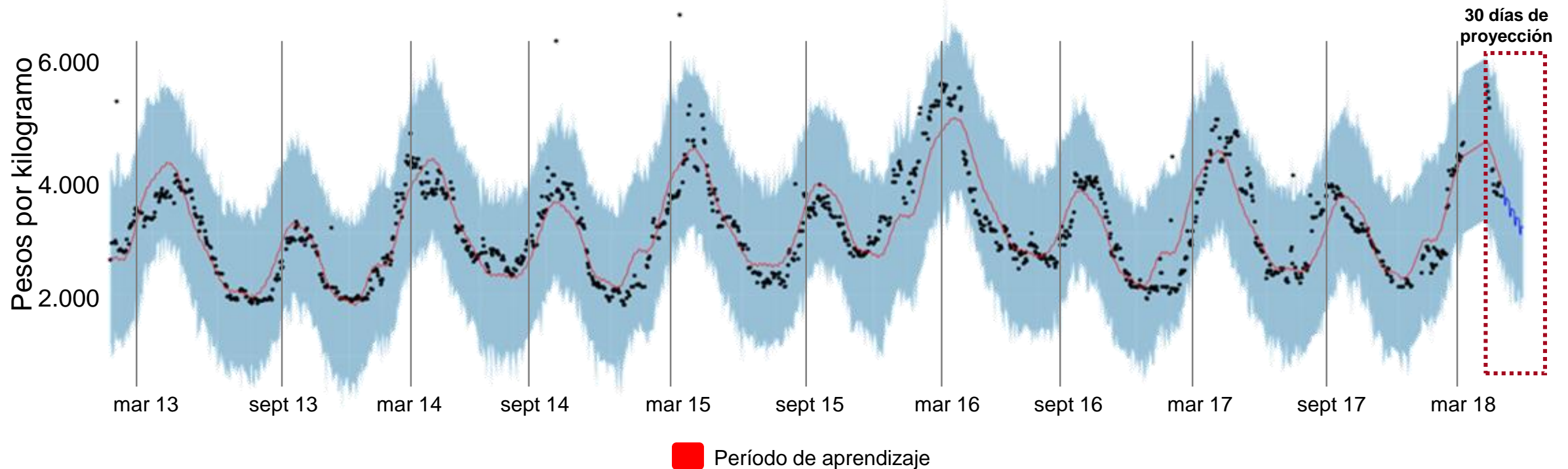
3.399.276 datos analizados

4. Big Data para el monitoreo de precios agropecuarios

Para un periodo de 30 días se anticipa el comportamiento del precio de la mandarina que muestra tendencia a la baja

Mandarina Arrayana – Central de Abastos de Bogotá (enero 2013 – abril 2018)

La precisión relativa del modelo por producto es de **80,3 %**



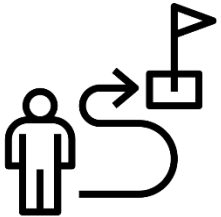
5. Big Data para predicción de dificultades en los proyectos del SGR

Módulo de inteligencia artificial para evaluar proyectos presentados al SGR



Objetivo

Detección anticipada de dificultades en la ejecución de los proyectos que permita focalizar la labor de seguimiento que realiza la Entidad



Metodología

Algoritmo de procesamiento natural del lenguaje

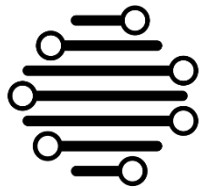
- Depuración de texto no estructurado de la formulación de proyectos

Uso de redes neurales

- Transformación del texto de los proyectos a información cuantificable representada mediante vectores

Algoritmo de clasificación “*support vector machine*”

- Determina la ocurrencia o no ocurrencia de dificultades en los proyectos



Insumos

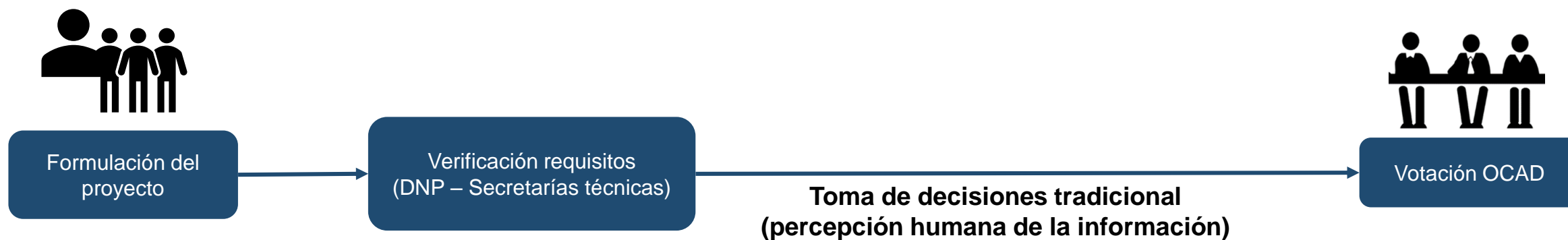
52.650 registros del Índice de Gestión de Proyectos de Regalías (IGPR)

10.530 documentos de formulación de proyectos

1.591.595 palabras depuradas de más de 250 millones de palabras totales (en la formulación de los proyectos)

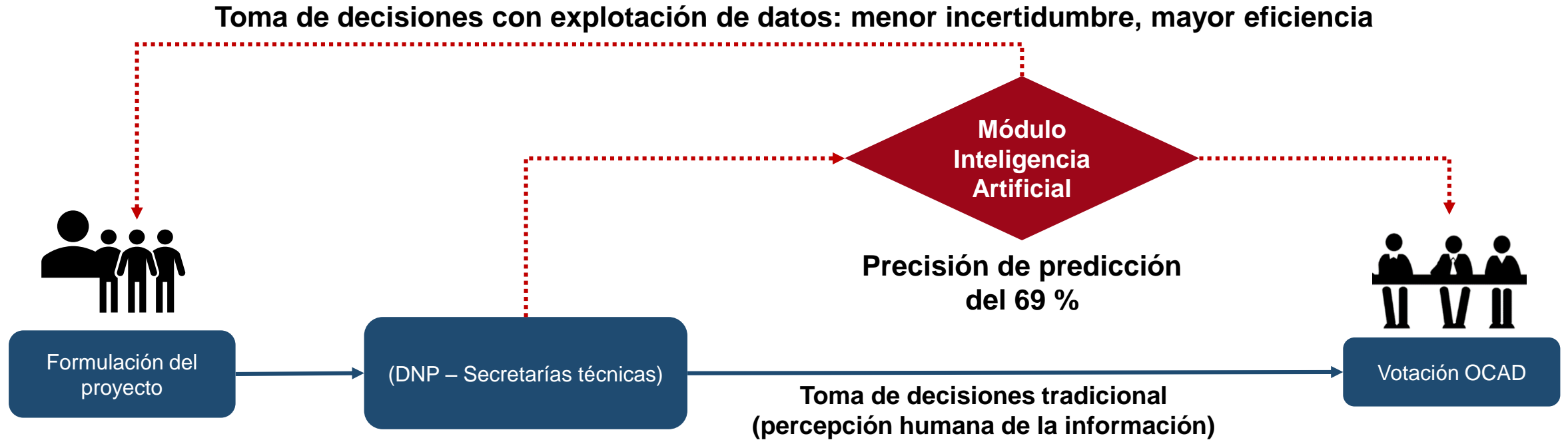
5. Predicción de dificultades en los proyectos del SGR

La aprobación de los proyectos se realiza con la información limitada a cada proyecto, disponible al momento de verificación de requisitos y votación



5. Predicción de dificultades en los proyectos del SGR

Reducción del riesgo en la ejecución de proyectos del SGR con una herramienta que fortalece la verificación de los proyectos previo a su aprobación





Departamento Nacional de Planeación
www.dnp.gov.co