

# Dirección de Desarrollo Digital

Unidad de Científicos  
de Datos



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



## VISOR DEL MODELO DE EXPLOTACIÓN DE DATOS

### INFORME FINAL

<b>Dependencias y entidades involucradas</b>	Departamento Nacional de Planeación <ul style="list-style-type: none"><li>• Dirección de Desarrollo Digital - Unidad de Científicos de Datos</li><li>• Consejería para Asuntos económicos y la transformación digital</li></ul>
<b>Sector</b>	TIC
<b>Tecnologías utilizadas</b>	R, Shiny
<b>Fuentes de datos</b>	Encuesta de Google Forms y datos de Google Sheets Tablero de shiny

## Contenido

<a href="#">1. <u>Presentación</u></a> .....	2
<a href="#">2. <u>Objetivos del proyecto</u></a> .....	2
<a href="#">3. <u>Metodología</u></a> .....	2
<a href="#">4. <u>Resultados</u></a> .....	7
<a href="#">5. <u>Conclusiones y recomendaciones</u></a> .....	9
<a href="#">6. <u>Socialización</u></a> .....	9
<a href="#">Contacto</a> .....	9



## Presentación

El modelo de explotación de datos es un instrumento conceptual estandarizado que tiene como objetivo fomentar el desarrollo de las actividades para el uso y explotación de datos en las entidades públicas. Este modelo toma en consideración las capacidades, los recursos tecnológicos, los recursos humanos y financieros, y los objetivos misionales de las entidades; entendiendo la importancia de la gestión y aprovechamiento de datos para el desarrollo de la economía digital y la adopción de tecnologías de la 4ª Revolución Industrial. El modelo de explotación de datos es una herramienta que tiene como fin contribuir a la mejora continua de las entidades, permitiéndoles gestionar y monitorear indicadores sobre el estado actual del aprovechamiento de sus datos y la hoja de ruta que deben seguir para alcanzar sus objetivos en esta materia.

*The data exploitation model is a standardized conceptual instrument. This model aims to develop the activities of use and exploitation of data in public entities. This model takes into account the capacities, technological resources, human and financial resources, and the mission objectives of the entities; understanding the importance of the management and use of data for the development of the digital economy and the adoption of technologies from the 4th Industrial Revolution. The data exploitation model is a tool that aims to contribute to the continuous improvement of entities, allowing them to manage and monitor indicators on the current state of the use of their data and the roadmap they must follow to achieve their objectives in this area.*

## Objetivos del proyecto

### 1.1. General

Desarrollar un aplicativo en Shiny a partir del tablero preexistente donde se implemente una nueva pestaña con el modelo de explotación de datos.

### 1.2. Específicos

1. Implementación del modelo de explotación de datos dentro de una herramienta de visualización.
2. Integrar los módulos de visualización de indicadores de aprovechamiento de datos al aplicativo existente a partir de los resultados obtenidos por el modelo de maduración.

## Metodología

El desarrollo del tablero para la visualización del modelo de explotación de datos involucra las siguientes etapas:

1. **Adecuación del aplicativo:** El aplicativo existente es un desarrollo en Shiny que contiene gráfica del avance de los proyectos relacionados con tecnología. A este tablero se le añadió una nueva pestaña denominada Modelo de maduración en donde se incluyó todo el desarrollo del proyecto.
2. **Formulario de aprovechamiento de datos:** Para el desarrollo del modelo, se definió un formulario con las preguntas e información requerida de parte de las entidades. En este se consolidaron una serie de preguntas relacionadas con las 6 dimensiones del modelo de madurez: Tecnológica, humana, financiera, organizacional, táctica y operativa. El formulario incluye preguntas para caracterizar a la entidad en relación con su tamaño, sector y presupuesto, así como preguntas que permiten conocer el nivel de madurez de las entidades en cada una de las dimensiones y el nivel de madurez que esperan tener a futuro. Con base en las respuestas del formulario se realiza el cálculo del modelo.
3. **Adaptación del modelo:** el modelo de explotación de datos incluye los siguientes puntos:
  - **Diagnóstico del nivel de madurez actual y nivel de madurez esperado:** La estimación del índice de madurez es un valor estandarizado entre 0 y 1 donde 0 es el nivel más bajo de madurez y 1 el nivel más alto. El primer paso para este cálculo, teniendo en cuenta que dentro del formulario las respuestas se encuentran como: Nivel de madurez, es asignarle un número a cada categoría, finalmente se tienen 5 categorías: números de 1 a 5. Una vez se realizó este procedimiento con la ecuación 1 se realiza el cálculo de nivel de madurez en cada una de las dimensiones.



Ecuación 1: cálculo del nivel de madurez

$$P_1(x_1, x_2, x_3, \dots, x_N)$$

$$P_1(z_1, z_2, z_3, \dots, z_N)$$

$$d(P_1, P_2) = \sqrt{(z_1 - x_1)^2 + (z_2 - x_2)^2 + (z_3 - x_3)^2 + \dots + (z_N - x_N)^2}$$

Fuente: Metodologías de estimación de los componentes del modelo de implementación de explotación de datos

$z_i$ : valor máximo que puede tomar la pregunta  $i$   
 $x_i$ : valor otorgado a la respuesta  $i$  de la entidad  
 $N$ : número de preguntas en la dimensión

Una vez realizado el cálculo de la Ecuación 1, se estandariza la medida obtenida con base en la distancia máxima posible, en este caso  $x_i$  pasa a ser 1(uno). Después de obtener ese valor se realiza el cálculo que se observa en la Ecuación 2.

Ecuación 2: estandarización

$$IM_j = 1 - \frac{d(P_j, P_{max})}{d(P_{min}, P_{max})} * 100$$

Fuente: Metodologías de estimación de los componentes del modelo de implementación de explotación de datos

La Ecuación 2 genera un número que debe ser evaluado teniendo en cuenta los criterios de la Ecuación 3.

Ecuación 3: clasificación según puntajes

$$Nivel = \left\{ \begin{array}{l} 1, si 0 < Puntaje < 20 \\ 2, si 20 \leq Puntaje < 40 \\ 3, si 40 \leq Puntaje < 60 \\ 4, si 60 \leq Puntaje < 80 \\ 5, si 80 \leq Puntaje < 100 \end{array} \right\}$$

Fuente: Metodologías de estimación de los componentes del modelo de implementación de explotación de datos

La Ecuación 3 da finalmente el nivel de madurez de las entidades en cada una de las dimensiones.

• **Estimación del valor potencial a partir de Big Data:**

Dentro del formulario se encuentran preguntas para las siguientes dimensiones que permiten conocer el presupuesto destinado por la entidad, así como el porcentaje estimado de valor público creado con dichas inversiones para las siguientes dimensiones:



- *Explotación de datos y Big Data*
- *Recurso humano destinado a la explotación de datos*
- *Infraestructura tecnológica*
- *Desarrollo de proyectos, prototipos y/o pruebas concepto*

Estos porcentajes son transformados en valores nominales, con lo cual se obtiene una estimación del valor público potencial monetizado que obtiene la entidad con base en cuatro marcos de referencia: la inversión en explotación de datos y Big Data, la inversión en recurso humano destinado a la explotación de datos, la inversión en infraestructura tecnológica y la inversión en proyectos, prototipos o pruebas de concepto.

Finalmente se calcula el índice de valor potencial para cada uno de los 4 marcos de referencia con base en el presupuesto general de la entidad para el año de análisis. Este se obtiene con el procedimiento descrito en la *Ecuación 4*

*Ecuación 4: cálculo índice de valor potencial*

$$\text{índice} = \frac{\text{valor potencial}_i}{\text{presupuesto total}}$$

*Fuente: elaboración propia*

- **Estimación de inversión requerida para avanzar de nivel de madurez y el retorno esperado dada la inversión:**

Para este paso se tiene como insumo:

1. El nivel de madurez al que la entidad desea llegar en la dimensión de recurso humano y tecnológico, el cual se obtiene de las respuestas del formulario.
2. El nivel de complejidad o riqueza de los datos de la entidad se clasifica en tres niveles: alto, medio y bajo, este se calcula con base en la variedad de las tipologías de datos que se manejan dentro de las entidades. El cálculo del nivel de complejidad se basa en dos secciones de preguntas que se encuentran en el formulario, la primera corresponde a la capacidad total de almacenamiento disponible, capacidad de uso y capacidad libre, estas tres medidas en giga bytes, distinguiendo almacenamiento de infraestructura propia de la entidad y en la nube y la segunda corresponde a las proporciones de datos (de la capacidad en uso) en las distintas tipologías que tiene almacenada y disponible la entidad en medios digitales.

Con esta información se asocia la complejidad y riqueza de los datos a las proporciones de los datos almacenados según tipología y se obtiene obtener una clasificación objetiva con el índice de entropía de Rényi para el caso especial de divergencia  $\alpha=1$ , es decir basados en la fórmula de Shannon que mide la incertidumbre de una fuente de información:

*Ecuación 5: Fórmula de Shannon*

$$\lambda = \frac{-\sum_{i=1}^R p_i * \log_2 p_i}{\log_2 R}$$

*Fuente: Metodologías de estimación de los componentes del modelo de implementación de explotación de datos*

$p_i$ : proporción de datos de la tipología  $i$

$R$ : número de tipologías de datos máximo que tiene la entidad.



Si  $\lambda < 0,3$  la entidad tiene un nivel de complejidad bajo, si  $\lambda > 0,7$  la entidad tiene un nivel de complejidad alto. Para el resto de los casos la entidad tiene un nivel de complejidad medio.

Teniendo en cuenta lo anterior, se construyeron 5 matrices de precios estándar para estimar la inversión requerida, en los recursos tecnológicos y humanos.

A continuación, se presentan las matrices usadas para el cálculo de inversión en recurso Tecnológico y recurso Humano:

Tabla 1: Precios de servicio, servidor para procesamiento

Procesamiento					
Nivel de Madurez	Nivel de complejidad			Unidad	Tiempo
	Alto	Medio	Bajo		
5	\$33.870.128	\$22.080.617	\$21.477.474	Unidad	Mes
4	\$16.750.275	\$13.353.016	\$11.376.811	Unidad	Mes
3	\$9.075.574	\$9.058.374	\$5.672.677	Unidad	Mes
2	\$6.152.921	\$6.504.712	\$3.987.046	Unidad	Mes
1	\$4.631.570	\$3.890.718	\$2.991.201	Unidad	Mes

Fuente: DNP, DDD

Tabla 2: Precios de servicio - base de datos Postgres

Software					
Nivel de Madurez	Nivel de complejidad			Unidad	Tiempo
	Alto	medio	Bajo		
5	\$18.482.455	\$14.002.836	\$12.828.472	Unidad	Mes
4	\$16.103.898	\$8.232.975	\$9.056.898	Unidad	Mes
3	\$11.184.551	\$6.516.839	\$6.241.611	Unidad	Mes
2	\$7.001.505	\$4.816.031	\$3.920.398	Unidad	Mes
1	\$3.001.160	\$2.408.416	\$1.826.885	Unidad	Mes

Fuente: DNP, DDD

Tabla 3: Precios de instalación servidor para procesamiento

Procesamiento				
Nivel de Madurez	Nivel de complejidad			Unidad
	Alto	Medio	Bajo	
5	\$3.715.729	\$4.297.095	\$4.293.825	Unidad
4	\$2.757.612	\$2.856.082	\$2.856.082	Unidad
3	\$2.310.439	\$1.880.826	\$1.880.826	Unidad
2	\$1.999.425	\$1.647.126	\$1.647.126	Unidad
1	\$1.722.862	\$1.438.766	\$1.438.766	Unidad

Fuente: DNP, DDD



Tabla 4. Precios de instalación - base de datos Postgres

Software				
Nivel de Madurez	Nivel de complejidad			Unidad
	Alto	Medio	Bajo	
5	\$5.589.089	\$4.810.999	\$4.810.999	Unidad
4	\$4.442.896	\$3.307.188	\$3.307.188	Unidad
3	\$3.880.819	\$2.962.788	\$2.918.180	Unidad
2	\$3.250.121	\$2.661.757	\$2.448.578	Unidad
1	\$2.600.882	\$2.219.348	\$2.172.930	Unidad

Fuente: DNP, DDD

Tabla 5. Salarios según nivel de complejidad y tipo de cargo

Nivel de complejidad	Salarios			
	Directivo	Asesor	Profesional	Técnico
Alto	\$13.012.575	\$10.606.242	\$7.078.107	\$2.751.526
Medio	\$8.821.666	\$7.412.475	\$4.842.289	\$1.953.089
Bajo	\$5.069.201	\$4.278.440	\$2.751.681	\$1.179.935

Fuente: DNP, DDD

Como se puede observar en la Tabla 5 es necesario determinar el número de funcionarios para cada una de las categorías (Directivo, asesor, profesional y técnico). Para ello se utiliza la información contenida en la Tabla 6.

Tabla 6. Número de empleados dedicados a la explotación de datos según categoría

Categorías	Número empleados requeridos
Técnicos	$\frac{M}{5}$
Profesionales	$\frac{M}{5}$
Asesores	$\frac{M}{20}$
Directivos	$\frac{M}{50}$

Fuente: DNP, DDD

La inversión requerida en capital humano se estima sumando los salarios anuales, correspondiente al número de funcionarios requeridos en cada una de las categorías (técnico, profesional, asesor y profesional).



Por otro lado, la inversión requerida en capital tecnológico se estima realizando la suma de los valores de inversión en almacenamiento y procesamiento, tanto de servicio como de instalación. Con respecto a la inversión de servicio en almacenamiento el cálculo se realiza multiplicando el precio de servicio anualizado (hallado con base en la tabla de servicio de almacenamiento) por la capacidad total de almacenamiento (en terabytes).

Con respecto a la inversión en instalación, esta se halla con base en la tabla de precios de instalación (tanto de procesamiento como de almacenamiento) y se multiplica por  $\frac{1}{5}$  con el objetivo de reconocer el valor de la depreciación de los activos tecnológicos.

La inversión requerida total anual se calcula sumando los valores de inversión en instalación de recurso tecnológico, servicio de recurso tecnológico y recurso humano.

- **Definición de una hoja de ruta para avanzar de nivel.**

Para cada uno de los puntos anteriores se realizan cálculos a partir de las respuestas de los usuarios en el formulario y como resultado el tablero de visualización muestra los siguientes conceptos para cada una de las dimensiones:

- Nivel de madurez actual, nivel de madurez esperado y recomendaciones para alcanzar el nivel de madurez esperado.
- Valor potencial e índice de valor potencial
- Inversión requerida

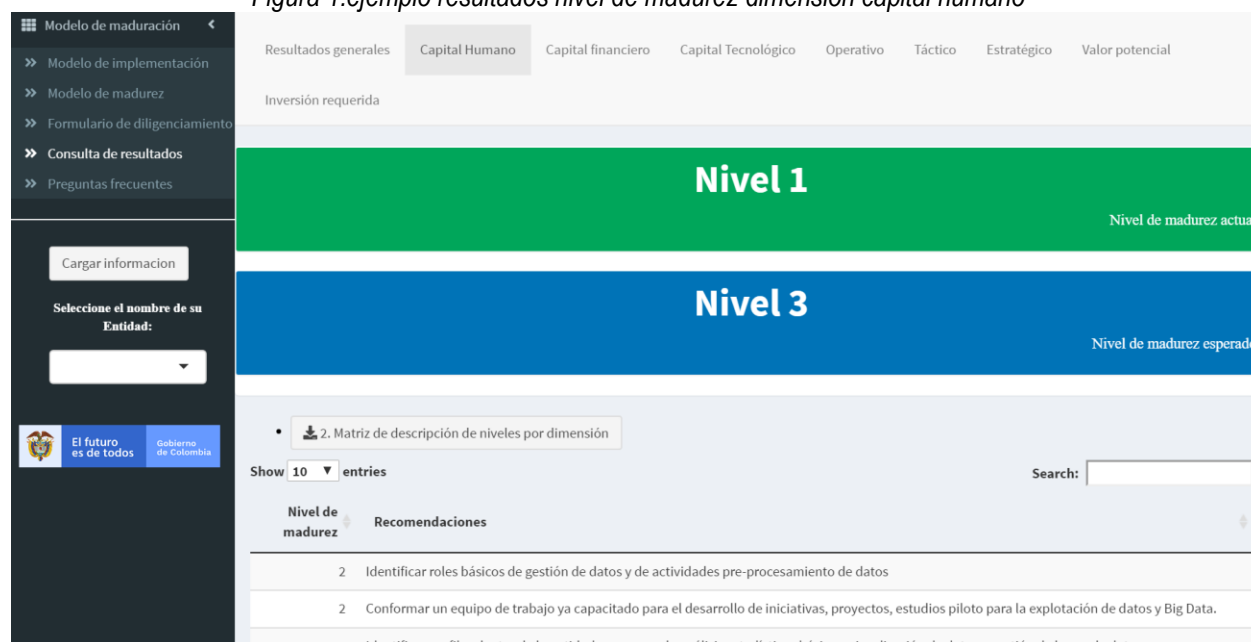
## Resultados

A través del desarrollo metodológico descrito en la Sección 0, se obtuvieron los resultados que se presentan a continuación. Toda retroalimentación desde un punto de vista experto o de usuario por parte de la Dirección de Desarrollo Digital y la Consejería para Asuntos económicos y la transformación digital es bienvenida. Este insumo será de gran ayuda para mejorar la calidad y utilidad de los resultados obtenidos, de manera que agreguen mayor valor.

1. **Tablero de visualización con la información consolidada:** Dentro del tablero preexistente de shiny con el fin de tener una herramienta útil y fácil de utilizar se incluye:
  - El formulario con una guía de diligenciamiento para las entidades. Una vez el usuario diligencia el formulario debe cargar la información.
  - Una vez es cargada la información del formulario al tablero se puede consultar el nivel de madurez en el que se encuentran las entidades para cada una de las dimensiones en el momento en que diligencian el formulario. Asimismo, se observa el nivel de madurez que quieren alcanzar y una serie de recomendaciones para alcanzar este nivel. Un ejemplo de esto se puede observar en la *Figura 1*.



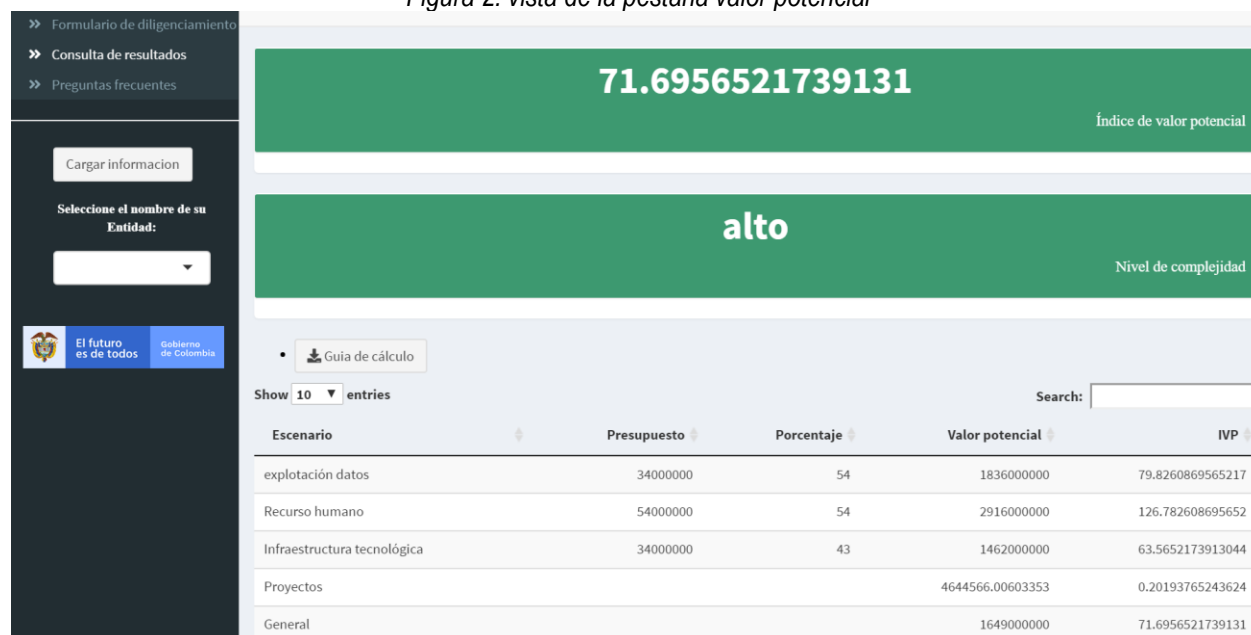
Figura 1: ejemplo resultados nivel de madurez dimensión capital humano



Fuente: elaboración propia

- Por otra parte, se encuentra el valor potencial de los proyectos desarrollados por las empresas, en esta pestaña se encuentra el valor potencial conjunto, el nivel de complejidad y una tabla con el valor potencial desglosado por los diferentes escenarios (explotación de datos, recurso humano, infraestructura tecnológica, proyectos y finalmente el general).

Figura 2: vista de la pestaña valor potencial



Fuente: elaboración propia



- Finalmente se encuentra la pestaña de inversión requerida que muestra la inversión requerida total anual y el valor desagregado por dimensión tecnológica y de recurso humano. En la *Figura 3* se observa un ejemplo de esta pestaña.

Figura 3: vista de la pestaña inversión



Fuente: elaboración propia

## Conclusiones

A partir de la metodología desarrollada y de los resultados obtenidos para cumplir los objetivos de este proyecto, planteados en el plan de trabajo, se presentan a continuación las principales conclusiones obtenidas por el equipo de la UCD y las principales recomendaciones para un mejor uso y aprovechamiento del proyecto.

1. El desarrollo del proyecto permite a las entidades conocer sus fortalezas y debilidades en actividades relacionadas con la explotación de datos, esto les permite planear rutas de mejora con el fin de sacar el mayor provecho de los datos y sus capacidades como entidad.
2. Con la información obtenida de las entidades se puede llevar un control de su progreso a lo largo del tiempo.
3. El tablero de visualización del modelo de explotación de datos permite tanto a las entidades como a la Consejería para Asuntos económicos y la transformación digital y la Dirección de Desarrollo Digital del DNP, conocer toda la información relevante para un mejor aprovechamiento de los datos.

## Socialización

Los resultados del proyecto han sido socializados durante su desarrollo con la Dirección de Desarrollo Digital y la Consejería para Asuntos económicos y la transformación digital.

## Contacto

Si tiene alguna duda, comentario o sugerencia sobre este proyecto, o si le gustaría conversar con la Unidad de Científicos de Datos sobre la posibilidad de una nueva fase para el mismo, puede comunicarse con nosotros a través del correo electrónico [ucd@dnp.gov.co](mailto:ucd@dnp.gov.co).