



El futuro  
es de todos

Consejería Presidencial  
para asuntos económicos  
y transformación digital



El futuro  
es de todos

DNP  
Departamento  
Nacional de Planeación



El futuro digital  
es de todos

MinTIC

# ANEXO 4

Metodologías de cálculo del modelo



## METODOLOGÍAS DE ESTIMACIÓN DE LOS COMPONENTES DEL MODELO DE IMPLEMENTACIÓN DE EXPLOTACIÓN DE DATOS

### 1. Estimación del nivel de madurez de explotación de datos

La estimación de un índice de madurez puede hacerse a través de varias técnicas y su definición depende del interés que haya sobre su construcción, es decir de lo que se espera de dicho índice. La aproximación que se tomó para el *modelo de madurez de explotación de datos* busca obtener un índice estandarizado entre 0 y 1, esto es que si  $IM$  es el índice de madurez de la entidad  $j$  puedes definirlo como un valor continuo tal que  $IM_j \in [0,1]$ ; donde, por ejemplo, 0 se asigne a una entidad hipotética —no tiene que existir en el conjunto de las entidades observables—, que tiene el nivel más bajo de madurez, y 1 se asigne a otra entidad hipotética que tiene el nivel más alto de madurez posible medido con los instrumentos diseñados para tal fin.

La alternativa inicial se basa en una distancia euclidiana, relativa al valor mínimo posible, del hiperplano que produce cada una de las  $N$  preguntas del instrumento asociado al modelo de madurez definido. Para esto, se generalizó el uso de la ecuación de distancia en un espacio con  $N$  dimensiones; donde  $N$  es definido como el número de preguntas. Antes de entrar en detalle sobre el

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

$$P_2(z_1, z_2, z_3, \dots, z_N)$$

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



procedimiento de cálculo, vale recordar la ecuación de distancia entre dos puntos arbitrarios  $P_1$  y  $P_2$ , tal que:

Aquí se obtendrá una distancia donde tal que  $d_{1,2} \in R$ .

Ahora bien, para el cálculo del nivel de madurez de la entidad  $j$ , cuyos valores de respuesta al instrumento son  $x_{j1}, \dots, x_{jk}, \dots, x_{jN}$  que corresponden a los niveles de madurez arrojados en las  $N$  respuestas de dicha entidad, donde:  $x_{jk} = 1, 2, 3, 4$  o  $5$ , según sea la respuesta dada a la pregunta  $K$ , en la escala ordinal diseñada para el ítem.

Sentado lo anterior, se pasará a estandarizar dicha medida con base en la distancia máxima posible que genera el hiperespacio de  $N$  dimensiones, que está dada por la distancia entre la entidad hipotética que es evaluada con el mínimo rango del nivel, es decir,  $P_{min}$  en cada uno de los  $N$  ítems y la entidad hipotética que es evaluada con el máximo en todas y cada una de las preguntas, es decir,  $P_{max}$ . La estandarización entonces se hará relativa a la

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

distancia de  $P_{min}$  y el  $IM_j$  quedará definido así:

Finalmente, si se desea puede retornarse a una escala categórica para asignarlo a una entidad en su nivel de madurez alcanzado teniendo en cuenta las siguientes condiciones.

$$Nivel = \begin{cases} 1, & \text{si } 0 < Puntaje < 20 \\ 2, & \text{si } 20 \leq Puntaje < 40 \\ 3, & \text{si } 40 \leq Puntaje < 60 \\ 4, & \text{si } 60 \leq Puntaje < 80 \\ 5, & \text{si } 80 \leq Puntaje < 100 \end{cases}$$



Dónde el puntaje de la entidad es simplemente:

$$Puntaje = IM_j \times 100$$

Así mismo, esta medición se podría generalizar a cada una de las dimensiones del modelo de madurez.

Para tener mayor claridad sobre el cálculo, se plantea el siguiente ejemplo que propone el cálculo del nivel de madurez de una entidad cualquiera para la **dimensión de recurso humano**. Al asumir que la dimensión de recurso humano cuenta con 5 preguntas,  $N$  será igual a 5; de modo que el proceso de cálculo para este ejemplo será el siguiente:

$$P_j(4, 3, 4, 2, 3)$$

$$P_{max}(5, 5, 5, 5, 5)$$

$$P_{min}(1, 1, 1, 1, 1)$$

$$d(P_j, P_{max}) = \sqrt{(5-4)^2 + (5-3)^2 + (5-4)^2 + (5-2)^2 + (5-3)^2}$$

$$d(P_j, P_{max}) = 4.36$$

$$d(P_{min}, P_{max}) = 8.94$$

$$IM_j = 1 - \frac{4.36}{8.94} = 0.5127$$

$$Puntaje = 51.27$$

Teniendo en cuenta que el puntaje obtenido se encuentra entre 40 y 60, la entidad tendrá un nivel de madurez asignado de 3. Para el cálculo del nivel de madurez de las otras seis dimensiones del modelo se procede con el mismo procedimiento. A continuación, se presentan los resultados de una entidad ficticia que incluye el ejemplo planteado.



Tabla 1  
Ejemplo de cálculo niveles de madurez para una entidad ficticia

	Puntaje	Nivel de madurez
Recurso financiero	24,48	2
Recurso humano	51,27	3
Recurso tecnológico	45,40	3
Operativo ciclo de datos	66,67	4
Táctico gestión de datos	54,17	3
Estratégico gobernanza de datos	45,80	3

Fuente: DNP - DDD.

### Aclaración sobre el cálculo de nivel de madurez del recurso financiero

Para el cálculo de nivel de madurez de la dimensión de recurso financiero se hace un ajuste a las entidades que respondan niveles de madurez superiores a 1 y no reporten ningún proyecto, estudio piloto o prueba de concepto relacionado con explotación de datos y *big data*. La corrección al puntaje se aplica con base en que el recurso financiero es una de las dimensiones más importantes en el contexto del modelo y debe dar cuenta de la realidad institucional de la entidad, soportado por iniciativas concretas relacionadas con explotación de datos y *big data*. A continuación, se presenta la ecuación de corrección del nivel de madurez, que aplica exclusivamente para la dimensión financiera.

$$P_{\text{corregido}} = (P * \delta) + (I * P) \times (1 - \delta)$$

Donde  $P$  es el puntaje de la dimensión financiera —calculado como se explica en la sección anterior—,  $I$  es la razón entre el presupuesto destinado a actividades de explotación de datos y *big data* y el presupuesto total de la



entidad para el año en vigencia. *I* es una variable indicadora que toma los valores 1 o 0, según la entidad cuenta, o no, con iniciativas o proyectos relacionados con explotación de datos y *big data*. Como se puede observar en la ecuación as entidades que cuenten con proyectos de explotación de datos no tendrán ninguna corrección, o ajuste. Sin embargo, las que no reporten ninguna iniciativa o proyecto, su puntaje se ajustará de acuerdo con la proporción de presupuesto destinado a actividades de explotación de datos.

## 2. Estimación del valor potencial y del índice de valor potencial

La estimación del potencial que tiene cada entidad para generar valor público a partir de la explotación de datos y *big data* se inicia con el análisis del presupuesto que destina la entidad a las fuentes de creación de valor en esas actividades:

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

Tabla 2

*Destinación de presupuesto destinado a fuentes de creación a partir de la explotación de datos*

Escenario	Presupuesto
Explotación de datos	\$ 2.127.662.673
Recurso humano	\$ 101.709.912
Infraestructura tecnológica	\$ 15.334.966.407
Proyectos	\$ 1.374.175.228

Fuente: DNP, DDD



Una vez los quienes responden el formulario identifican el presupuesto asignado a las fuentes de creación de valor, se les solicita que estimen en qué medida esa inversión afecta o retribuye a los grupos de interés de la entidad y a la creación de valor público con base en su misionalidad. En este orden de ideas, se recolecta información sobre el porcentaje estimado de valor público que se genera a partir de la inversión en cuatro áreas:

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

Posteriormente, se transforman esos porcentajes en un valor nominal teniendo en cuenta el monto de la inversión indicado por la entidad en el formulario para estos rubros. Así, se estima el 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.

Por ejemplo, una entidad identifica que de los \$15.000 millones invertidos en infraestructura tecnológica para la explotación de datos, esto retribuye en un 95% la generación de valor público. Nominalmente eso representa una estimación de \$14.000 millones en generación de valor público a partir de la explotación de datos.



Tabla 3  
Destinación de presupuesto destinado a fuentes de creación a partir de la explotación de datos

Escenario	Presupuesto	Porcentaje de valor público	Valor Potencial
Explotación de Datos	\$ 2.127.662.673	0,9	\$1.914.896.405
Recurso Humano	\$ 101.709.912	0,975	\$99.167.164
Infraestructura Tecnológica	\$ 15.334.966.407	0,95	\$14.568.218.086
Proyectos	\$ 1.374.175.228	1	\$1.374.175.227
General			\$1.644.535.816

Fuente: DNP - DDD.

Para el caso particular de los proyectos, los prototipos y/o las pruebas concepto se hace un ajuste de tal forma que los valores públicos potenciales monetizados de cada proyecto se presenten con base anual para el análisis. Por ejemplo, si una entidad reporta estar implementando dos proyectos con una duración de tres años cada uno, el porcentaje monetizado de cada proyecto se divide por tres. Además, se suma el valor anual monetizado de los dos proyectos para construir el valor público potencial monetizado definitivo para el marco de proyectos.

## 2.1 Índice de valor potencial

Una vez calculado el valor potencial (monetizado) se contrastan los valores monetarios con el presupuesto total asignado para la entidad en la vigencia del año correspondiente a la evaluación. Este contraste da como resultado cuatro indicadores que representan el porcentaje del presupuesto que correspondería al valor público generado por las inversiones en explotación de datos y *big data*, en recurso humano destinado a la explotación de datos, en infraestructura



tecnológica y en desarrollo de proyectos, prototipos y/o pruebas concepto. Este indicador se denomina *índice de valor potencial* para cada fuente de creación de valor.

Para calcular el índice de valor potencial general se procede a eliminar los valores extremos<sup>1</sup> y se calcula la media restringida de los dos valores restantes, que se asigna como índice de valor potencial general. Este enfoque permite crear tanto un intervalo como un único índice de valor potencial que permitirá efectuar una comparación más estandarizada entre las entidades y ubicarlas en un *ranking* que les permita contrastar su nivel de madurez con el potencial en términos de explotación de datos y *big data* que se evidencia a la fecha.

---

<sup>1</sup> Al eliminar los valores extremos de los marcos de análisis es posible limpiar el indicador final de posibles subestimaciones o sobreestimaciones del valor público potencial que generan ciertas inversiones.