



El futuro digital
es de todos

Gobierno
de Colombia
MinTIC

Identificación, análisis y
correlación de brechas de
información pública y
recomendaciones



Ministerio de Tecnologías de la Información y las Comunicaciones
Viceministerio de Economía Digital
Dirección de Gobierno Digital

Equipo de trabajo MinTIC - Datos Abiertos

Luisa Fernanda Medina
Carolina Matamoros
Carlos Julio León
Ana Carolina Escobar

Equipo de trabajo DNP – Planeación Dirección de Desarrollo Digital

Diana Paola Ramírez

Versión	Observaciones
Versión 1 Febrero 2020	Incluye el análisis y recomendaciones derivadas de la correlación conforme los compromisos del CONPES 3920

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

Metodología y modelo de uso de Datos Abiertos



Esta Guía para el uso y aprovechamiento de Datos Abiertos en Colombia de la Dirección de Gobierno Digital se encuentra bajo una [Licencia Creative Commons Atribución 4.0 Internacional \(Commons\)](https://creativecommons.org/licenses/by/4.0/).



Tabla de contenido

Tabla de Contenido.....	¡Error! Marcador no definido.
1. Introducción	4
2. Desarrollo de la correlación	4
2.1. <i>Resultados del Índice de Gestión de Activos.....</i>	<i>5</i>
2.2. <i>Resultados del Índice de Madurez.....</i>	<i>6</i>
2.3. <i>Desarrollo de la Correlación</i>	<i>9</i>
2.4. <i>Resultados Correlación.....</i>	<i>11</i>
3. Análisis y Recomendaciones	13



1. Introducción

En el desarrollo del CONPES 3920 de Explotación de Datos, en de su cuarto objetivo específico “Generar cultura de datos en el país” aclara que el mismo es “un elemento habilitador del consumo” de datos, así mismo definen que para los temas de interés de la política básicamente se pueden clasificar a los usuarios de datos en dos categorías distintas, una de ellas es la ciudadanía, la cual en el pleno desarrollo de sus derechos al hacer uso de los datos abiertos ejerce su capacidad de veeduría y supervisión sobre los asuntos del estado. Sin embargo, la segunda categoría es no sólo la que corresponde el interés de este informe, sino aquella que permite que el Estado y sus entidades contar con la capacidad de producir datos de calidad y además utilizarlos.

Esta categoría corresponde con la de los tomadores de decisiones basadas en datos, puntualmente el estudio busca revisar a partir de los resultados tanto del índice de gestión de activos de información como el índice de Madurez de Big Data. Estas dos fuentes de información adicionalmente también fueron desarrolladas en el marco del CONPES 3920 buscando generar un diagnóstico del estado actual de procesamiento y madurez de las entidades públicas de su información.

Como hipótesis se plantea que a mayor índice de gestión de activos de información mayor sería la madurez en Big Data de las entidades. Esto presupone que las entidades que efectivamente publican de manera accesible, estructurada, frecuente, entre otros criterios del índice de gestión de activos efectivamente tienen más probabilidad de contar los elementos organizacionales, tecnológicos, jurídicos, culturales y éticos para poner en marcha o desarrollar procesos de Big Data dentro de su operación recurrente.

En el documento se analizan los resultados de la correlación realizada entre ambos ejercicios y finaliza realizando recomendaciones futuras, tanto para fortalecer este tipo de diagnóstico como a las entidades para fortalecerse de la forma más costo eficiente y rápida posible de forma que puedan mejorar su gestión de datos para el uso de Big Data en sus procedimientos

2. Desarrollo de la correlación

Cómo se mencionó brevemente en la introducción este ejercicio busca probar la hipótesis de que el índice de gestión de activos de información está correlacionado con el índice de madurez de Big Data en las entidades públicas.

De tal forma este ejercicio parte de los resultados obtenidos en el “Modelo analítico para la identificación de brechas de acuerdo con los registros de activos de información publicados en el portal de datos abiertos” realizado en 2018 en el Ministerio de Tecnologías de la



Información y las Comunicaciones, el cual finaliza con el desarrollo del índice de gestión de activos de información.

Así mismo este ejercicio parte de los resultados obtenidos en el “Índice de madurez de Big Data” realizado en 2019 por el Departamento Nacional de Planeación, en cual finaliza con el índice general de madurez para las entidades participantes en ese ejercicio de autoevaluación al igual que el subíndice de madurez en cinco grandes categorías, Organizacional, Cultural, Tecnológica, Jurídica y Ética.

2.1. Resultados del Índice de Gestión de Activos

Este ejercicio terminó un aplicativo disponible [aquí](#)¹ en este para inicios de 2020 se identifica que 146 entidades están siendo evaluadas y que en promedio el índice de madurez recibe una calificación de 26.4%. Adicionalmente, se muestra un análisis que cruza completitud con facilidad² en donde se ubican las entidades conforme al desempeño de esas dos variables clave. Con esto en mente de las 146 entidades diagnosticadas sólo 8 logran ubicarse en el primer cuadrante (de alta completitud y alta facilidad).

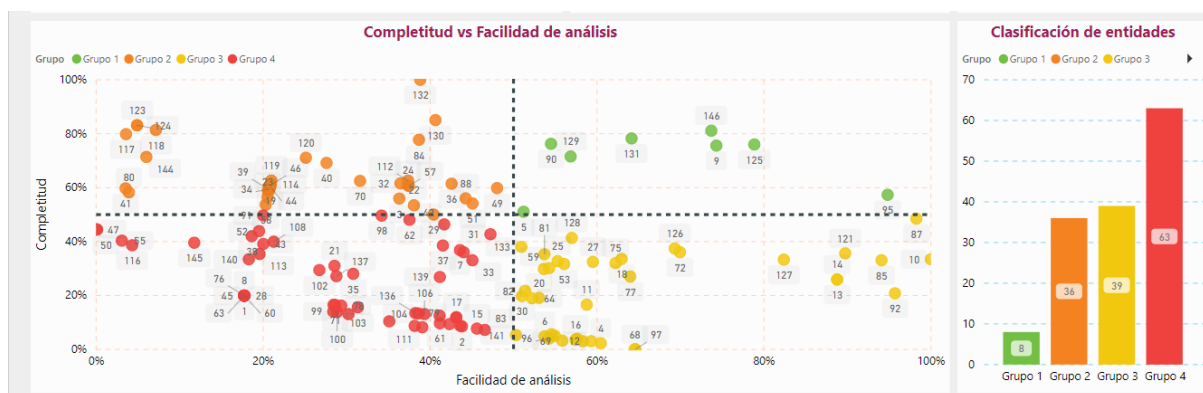


Ilustración 1 Valor del índice por Entidad: Fuente [Herramienta de Análisis del Índice de Gestión de Activos, MinTIC](#)

En este mismo tablero además puede encontrarse el conjunto de datos que muestra el resultado por entidad y para cada una de las variables del estudio. Así mismo este conjunto puede ser consultado en datos abiertos en <https://www.datos.gov.co/dataset/Activos-de-Informacion-de-Colombia/5cng-ebxj>

¹<https://app.powerbi.com/view?r=eyJrIjojNTA3ODZmNWItZmZmYS00Yjc2LWE3OWQtMWE3NjBiZjk1MTk1IiwidCI6IjRhNDQ5OTAzLWNiYTktNDk3ZS1hNzI4LWNlMzUzZjA2M2JkOSIsImMiOiJF9>

² Estas son dos variables compuestas que recogen todas las calificaciones realizadas en el estudio así:

- Completitud: compuesto por las variables calificación, tipo excepción y fundamentos.
- Facilidad: compuesto por las variables electrónico, formato y frecuencia de actualización.



Id	Grupo	Nombre	Clasificación de información	Formato	Frecuencia de actualización	Fundamento constitucional o legal	Idioma	Excepción parcial o total	Medio de conservación y/o soporte	Información disponible y/o publicada	URL de consulta	Plazo de la clasificación o reserva	Complejidad	Facilidad análisis	Índice madurez
		Armero													
96	3	Administradora Colombiana De Pensiones Colpensiones	0,0 %	84,8 %	0,0 %	0,0 %	0,0 %	0,0 %	67,6 %	0,0 %	0,0 %		4,0 %	54,3 %	9,2 %
109	4	Agencia De Desarrollo Local De Itagüí	0,0 %	56,6 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	61,4 %	94,5 %		13,1 %	30,2 %	7,8 %
77	3	Agencia De Renovación Del Territorio	0,0 %	99,3 %	0,0 %	60,7 %	0,0 %	60,7 %	84,8 %	0,0 %	91,7 %		27,0 %	64,0 %	31,3 %
60	4	Agencia Nacional De Defensa Jurídica Del Estado	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	0,0 %	61,4 %	76,6 %		19,9 %	17,7 %	6,8 %
87	3	Agencia Nacional De Minería Anm	0,0 %	95,2 %	91,7 %	56,6 %	0,0 %	66,9 %	80,0 %	56,6 %	64,8 %		48,5 %	98,3 %	55,8 %
92	3	Agencia Nacional Inmobiliaria Virgilio	0,0 %	96,6 %	89,0 %	0,0 %	0,0 %	0,0 %	80,7 %	0,0 %	0,0 %		20,8 %	95,7 %	30,7 %

Ilustración 2 Conjunto de datos de resultados del estudio del índice de gestión de activos [Herramienta de Análisis del Índice de Gestión de Activos, MinTIC](#)

Es con estos resultados que se procede a realizar la correlación.

2.2. Resultados del Índice de Madurez

En cuanto al índice de Madurez de Big Data en entidades públicas este se realizó mediante un estudio selectivo de autoevaluación en donde a las entidades le realizaban cuestionarios sobre las 5 dimensiones del análisis que ya han sido mencionadas.

Posteriormente los resultados eran agregados de forma ponderada de forma que cada dimensión obtenía su calificación, finalmente para obtener el índice de madurez el estudio realizó tres cálculos relevantes para la calificación global, estos eran: Criterio de experto, Ponderaciones Balanceadas³

A diferencia de los resultados del índice anterior, en donde participan todas las entidades que publican sus activos de información en el portal de datos abiertos, en este estudio se hizo una preselección de entidades que a) estuvieran de acuerdo con participar en el estudio b) tuvieran el tiempo y la asignación de personal para participar en el estudio. Por este motivo la muestra de este estudio se limitó a evaluar únicamente a 31 entidades.

Así pues, cada una de las dimensiones fue evaluada cada una con consideraciones críticas basadas en la revisión de literatura de la siguiente forma:

³ Es decir cada dimensión representa un peso equivalente en el estudio.

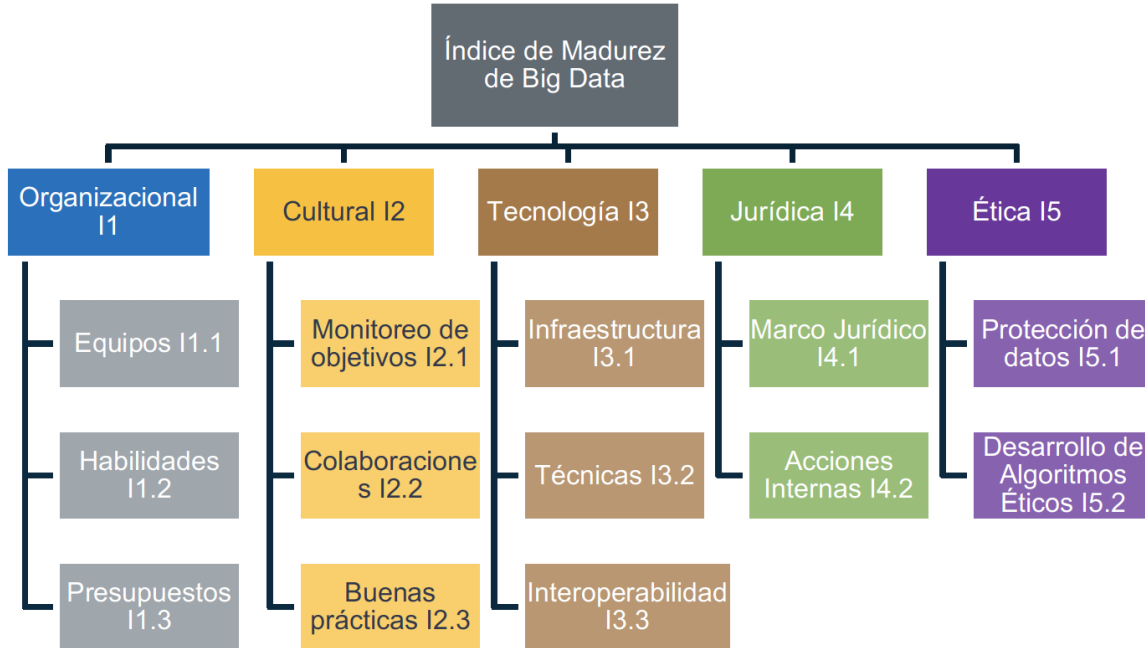


Ilustración 3: Dimensiones evaluadas en el estudio. Fuente: DNP

De esta forma por ejemplo para evaluar Organizacional se consideraba la conformación de los equipos, las habilidades TIC y el presupuesto asignado para el desarrollo de sus labores. Esta herramienta de diagnóstico puede ser consultada en línea [aquí](#).⁴

Así mismo los resultados pueden ser consultados en línea en la página web en la [sección de resultados](#)⁵, aquí puede consultarse por entidad al igual que de forma agregada por dimensión.

⁴ <https://bigdata.galileoiys.es/herramienta/>

⁵ <https://bigdata.galileoiys.es/resultados-2019/>



Fiscalía General de la Nación: (Colombia)

60,2 %
Desarrollo intermedio

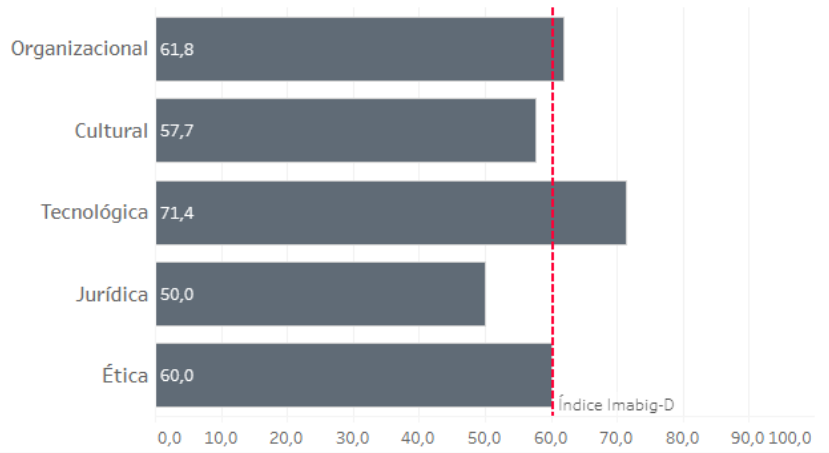


Ilustración 4 Resultado para una entidad. Fuente: <https://bigdata.galileoiys.es/resultados-2019/>

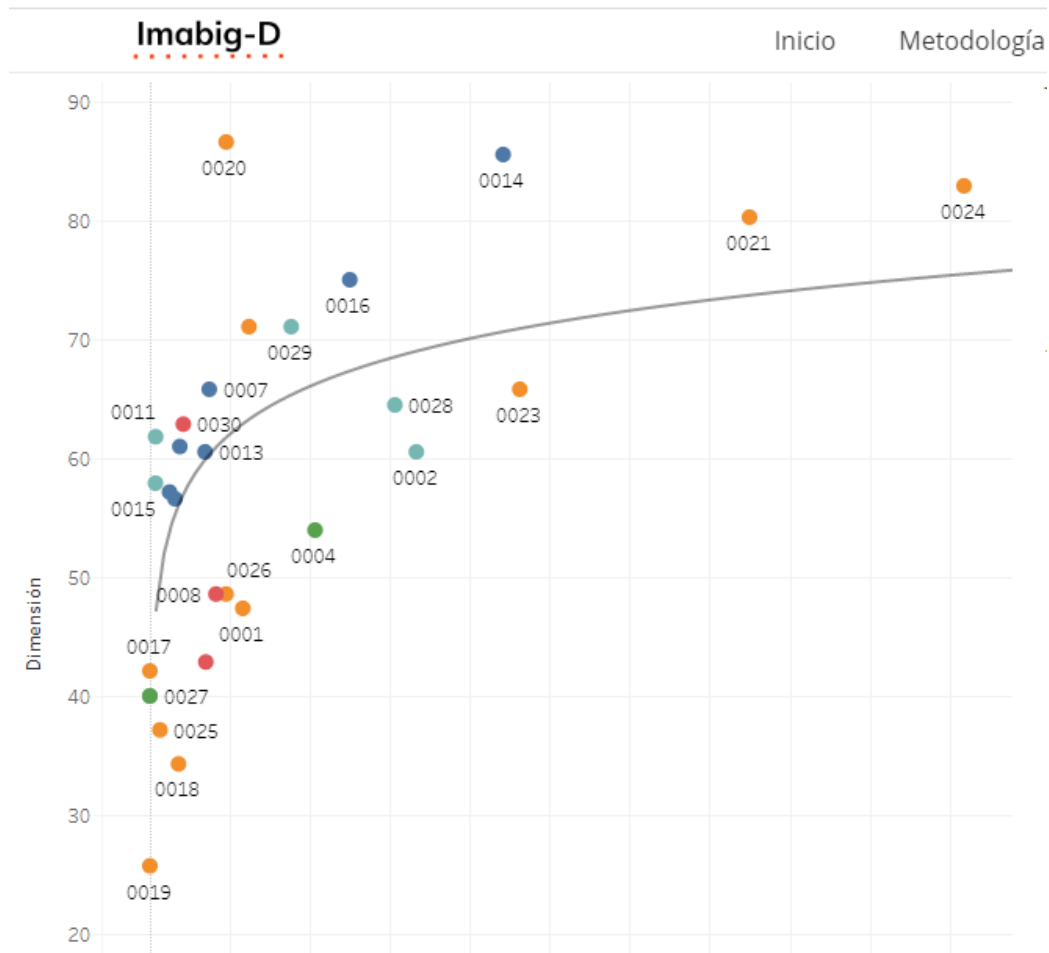


Ilustración 5 Resultados para la dimensión Organizacional. Fuente: <https://bigdata.galileiivs.es/resultados-2019/>

Así mismo y ya para manejo interno de información del Departamento Nacional de Planeación los resultados fueron agregados en un conjunto de datos que fue compartido con el grupo de datos abiertos del Ministerio de Tecnologías de la Información y Comunicaciones para desarrollar este estudio.

2.3. Desarrollo de la Correlación

Una vez se contó tanto con la base de datos del índice de gestión de activos de información como con el índice de madurez de Big Data se procede a realizar una correlación simple entre ambos resultados.

No obstante, para poder realizar este procedimiento nos percatamos de varios inconvenientes, el primero de ellos era el de manejo de los nombres de entidades y de la información. Por esto la primera necesidad se derivó en estandarizar los nombres de las entidades que aparecen en la herramienta de índice de gestión de activos de información. Por esta razón actualmente todas las entidades aparecen con formato de nombre propio y en formato UTF-8.



Posterior a esto fue necesario verificar qué entidades aparecían en ambos estudios. El resultado fue de tan sólo 14 entidades. Creando la principal limitante de este estudio, con esa cantidad de observaciones los resultados no son estadísticamente significativos, si bien pueden ser empleados a modo de indicativo no son una cantidad suficiente de observaciones como para definir si efectivamente la gestión de activos y la madurez de Big Data por parte de las entidades tiene una correlación.

En todo caso con esas 14 observaciones se procedió a realizar una correlación simple y el cálculo del coeficiente R², el cual también ayuda a la confiabilidad de que el comportamiento de una variable dependiente se explique por el comportamiento de la otra variable independiente, esto último implica el planteamiento de una hipótesis de causalidad. En esta oportunidad el ejercicio se realizó bajo el supuesto de que a mayor índice de gestión de activos mayor índice de madurez, en otras palabras, se tomó como variable dependiente el índice de madurez.

Estas son las entidades que pudieron participar en la evaluación:

1. Instituto Colombiano Para La Evaluación De La Educación
2. Instituto Nacional De Salud
3. Ministerio De Tecnologías De La Información Y Las Comunicaciones
4. Superintendencia De Industria Y Comercio
5. Servicio Geológico Colombiano
6. Fiscalía General De La Nación
7. Ministerio De Justicia Y Del Derecho
8. Ministerio De Agricultura Y Desarrollo Rural
9. Instituto Colombiano De Bienestar Familiar
10. Comisión De Regulación De Agua Potable Y Saneamiento Básico
11. Colpensiones - Administradora Colombiana de Pensiones
12. Departamento Nacional De Estadística
13. Comisión De Regulación De Energía Y Gas
14. Superintendencia Nacional De Salud

Así mismo se realizó el cálculo de ambos estadísticos tanto para los dos índices globales como para el índice de gestión de activos respecto a cada una de las dimensiones del índice de Madurez en Big Data.



2.4. Resultados Correlación

A partir de las consideraciones descritas en la subsección 2.3 los siguientes son los resultados obtenidos:

ENTIDAD	ORGANIZACIONAL	CULTURAL	TECNOLOGÍA	JURÍDICA	ÉTICA	INDICE MADUREZ
Instituto Colombiano Para La Evaluación De La Educación	85,5	100,0	73,8	65,0	84,2	81,7
Instituto Nacional De Salud	75,0	86,5	71,4	55,0	89,5	75,5
Ministerio De Tecnologías De La Información Y Las Comunicaciones	82,9	84,6	57,1	55,0	50,0	65,9
Superintendencia De Industria Y Comercio	62,9	59,6	64,3	75,0	41,2	60,6
Servicio Geológico Colombiano	64,5	92,3	50,0	55,0	40,0	60,4
Fiscalía General De La Nación	61,8	57,7	71,4	50,0	60,0	60,2
Ministerio De Justicia Y Del Derecho	80,3	65,4	52,4	60,0	29,4	57,5
Ministerio De Agricultura Y Desarrollo Rural	42,1	88,5	59,5	50,0	40,0	56,0
Instituto Colombiano De Bienestar Familiar	60,5	59,6	61,9	50,0	47,4	55,9
Comisión De Regulación De Agua Potable Y Saneamiento Básico	48,6	88,5	52,4	45,0	40,0	54,9

Colpensiones - Administradora Colombiana de Pensiones	65,8	63,5	47,6	45,0	29,4	50,3
Departamento Nacional De Estadística	58,6	36,5	47,6	50,0	47,1	48,0
Comisión De Regulación De Energía Y Gas	47,4	59,6	52,4	50,0	30,0	47,9
Superintendencia Nacional De Salud	42,9	53,8	38,1	45,0	40,0	44,0
Correlación (Pearson & Coef.de.correlación)	0,57	0,27	0,09	0,12	0,47	0,46
Coefficiente R2	0,33	0,07	0,01	0,01	0,22	0,21

Por tanto, se identifica que las dimensiones que más correlación tienen con el índice de activos son en su orden:

1. Organizacional
2. Ética
3. Cultural
4. Jurídica
5. Tecnológica

En todo caso en ninguna oportunidad la correlación es superior al 70%, en el índice agregado el resultado es tan sólo del 46% y para todos la posible explicación del índice de madurez a través del índice de gestión de activos está lejos de ser significativo (sobre 0.95) de hecho en todos los casos el bajo coeficiente de correlación indica con claridad que está haciendo falta mucha información para explicar las variables dependientes. Todo lo anterior bajo la fuerte influencia de sólo contar con 14 observaciones.



3. Análisis y Recomendaciones

Como se declaró con claridad estos no son significativos si bien dan una primera indicación en donde las dimensiones que más se correlacionan con el índice de gestión de activos son justamente las dimensiones “blandas” u “organizacionales”, es decir que las entidades podrían buscar enfocar sus esfuerzos más que en tecnología o en cambiar sus normativas en fortalecer su capital humano, sus equipos y procedimientos para efectivamente fortalecer su capacidad de hacer Big Data.

Esto está alineado fuertemente con las recomendaciones de los expertos que destacan la importancia del capital humano en las organizaciones cuando quieran embarcarse en la apropiación de una nueva tecnología.

Sin embargo, el estudio deja lecciones claras para poder fortalecer este diagnóstico en primera medida se requiere emplear los códigos DANE para identificar a las entidades, esto evita reprocesos, por ortografía o errores humanos al referirse a las entidades.

Así mismo la necesidad de alinear las iniciativas de las entidades antes de su desarrollo se hace manifiesta, por ejemplo, el estudio realizado en el Departamento Nacional de Planeación debió partir de la muestra original de los resultados del índice de gestión de activos, así se habría garantizado que las entidades participantes (sin importar el tamaño de la muestra) correspondieran con las entidades sobre las que se hizo el estudio de madurez.

Por último, se insta a las entidades a fortalecer estos estudios antes de proceder a realizar un estudio derivado de los mismos para garantizar que cualquier análisis derivado del mismo sí pudiera emplearse en la toma de decisiones a nivel nación.