

# Validación de Métricas propuestas de Calidad para el estudio de datos abiertos en base a criterios estándares: Aplicación HEVDA

Roxana Martínez<sup>1</sup>, Claudia Pons<sup>1,2,3</sup>, Rocío Rodríguez<sup>1</sup>, Pablo Vera<sup>1</sup>

<sup>1</sup> Universidad Abierta Interamericana (UAI)  
CAETI – Avenida Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina  
{Roxana.Martinez, Claudia.Pons, RocioAndrea.Rodriguez, PabloMartin.Vera}@uai.edu.ar

<sup>2</sup> LIFIA, Facultad de Informática, Universidad Nacional de La Plata (UNLP)  
Av. 7 n° 776, B1900 La Plata, Provincia de Buenos Aires, Argentina  
cpons@lifia.info.unlp.edu.ar

<sup>3</sup> Comisión de Investigaciones Científicas de la Provincia de Buenos Aires  
Calle 526 entre 10 y 11, B1900 La Plata, Provincia de Buenos Aires, Argentina

**Resumen.** Este trabajo presenta la metodología utilizada para la validación de métricas de calidad de datasets abiertos gubernamentales. Estas son sugeridas en base a los distintos aspectos definidos sobre estudios y relevamientos realizados. Para ello se explica cada métrica propuesta a través de un prototipo de desarrollo propio llamado HEVDA, el cual realiza un estudio de un dataset y verifica el estado de éste para cada una de las métricas sugeridas (registros repetidos, cantidad de casos con registros incompletos, entre otras), obteniendo así, un “estado de salud de los datos abiertos disponibilizados”. Además, se plantean los criterios estándares y dimensiones de calidad que son utilizados para la justificación de dichas métricas basadas en Normas ISO, buenas prácticas y estudios de criterios estándares universales en aspectos de calidad.

**Palabras Claves:** *Datos abiertos, Datos públicos, Calidad en datasets, Gobierno Abierto.*

## 1 Introducción

El papel de la tecnología es fundamental para promover el acceso a la información, la colaboración ciudadana y la disponibilidad de los datos. Es importante considerar ciertos facilitadores imprescindibles para un adecuado acuerdo de la utilización de estos aspectos entre el gobierno y la ciudadanía.

El enfoque principal de Gobierno Abierto se centra en la colaboración y participación de los ciudadanos, “la apertura del Estado y la agilidad en el diseño y

prototipado de políticas públicas, buscan que las iniciativas sean más abiertas, digitales, centradas en el usuario, basadas en evidencia y enfocadas en el futuro” [1]. Las herramientas tecnológicas que apoyan el marco de Gobierno Abierto permiten “promover y facilitar su uso siendo un instrumento que apoya el cumplimiento de los tres pilares del gobierno abierto: Transparencia, Participación y Colaboración” [2]. Además, “permite una mayor transparencia, ofrece servicios públicos más eficientes, y alienta un mayor uso público y comercial, y la reutilización de la información del gobierno. Algunos gobiernos incluso han creado catálogos o portales (como data.gov) para facilitar al público la búsqueda y el uso de esta información” [3].

El concepto de la apertura de datos públicos permite que la información de las entidades gubernamentales se presente a los ciudadanos en un formato abierto para que sea reutilizado por otros, es decir, aplicar el concepto de RISIP (Reutilización de la Información del Sector Público). Si bien el Estado Nacional administra grandes cantidades de datos públicos que son propiedad de todos los ciudadanos, y puede abrir estos datos y facilitar su acceso, es importante aclarar que esto ocurrirá siempre y cuando no exponga ningún tipo de información confidencial o privada, por lo que es necesario un adecuado análisis de los conjuntos de datos que serán considerados públicos y abiertos.

### **1.1 Gobierno Abierto y la inclusión de los datos abiertos**

La función principal de los datos abiertos gubernamentales es que sean reutilizados, por lo que definir un protocolo de estandarización en el proceso de apertura de datos, permite utilizar y reutilizar dicha información para que éstos sean interoperables por diversos organismos estatales, tanto a nivel nacional como internacional. “Los portales donde se publican datos abiertos son una oportunidad para democratizar el acceso a la información generada por las instituciones del sector público, así como potenciar su reutilización. Esta reutilización de datos abiertos permite el desarrollo de nuevos productos y servicios digitales, creando oportunidades de desarrollo social y económico” [4].

Es importante realizar un tratamiento sobre la calidad del dato brindado, ya que existen problemas que pueden ser mitigados con anterioridad, además se debe analizar hasta qué punto los datos sirven para los fines que fueron disponibilizados en los portales abiertos. “Dado que la calidad siempre es compleja de evaluar, ya que puede abordarse desde diferentes perspectivas y con diversas soluciones, es necesario definir lo que se entiende por calidad, como vamos a evaluar esa calidad y que nivel de calidad deseamos alcanzar” [5].

El Portal Europeo de Datos [6] es una iniciativa de la Comisión Europea, que recopila metadatos a partir de información del sector público disponible en portales de datos de acceso público de los distintos países europeos. También ofrece información sobre el proceso de suministro de datos y los beneficios de reutilizarlos. En uno de sus informes [7], publica que a medida que maduran las propuestas de datos abiertos de los países de la Unión Europea, su enfoque pasa de la cantidad a la calidad. Es decir, indica que, en los primeros años, menciona que se hicieron grandes esfuerzos para

publicar más y más datos, liberando datos activos que ya estaban en manos de las administraciones públicas. Sin embargo, sólo hubo unos pocos organismos estatales, que realizaron adaptaciones para mantener una calidad en los datos. En su artículo, también menciona que hay un enfoque intensificado en asegurar el valor de los datos para los reutilizadores, y que esto a menudo significa mejorar primero la calidad de los datos. Por lo anteriormente dicho, esto lleva a pensar que la importancia en abrir datos no es buscar meramente la disponibilización de grandes cantidades de datasets, sino en enfocarse en la calidad en el dato abierto para que pueda ser manipulada en múltiples plataformas tecnológicas.

El equipo de trabajo “Total Data Quality Management Program” [8] liderado por el profesor Richard Y. Wang del Instituto de Tecnología de Massachusetts (MIT University), definieron un conjunto de atributos y/o dimensiones para medir y gestionar la calidad de los datos que enfocadas en la evaluación que puede ser automatizada para valorar la idoneidad y adecuación de los datos en orden a objetivos de negocio o bien necesidades. Estudios posteriores han ido modificando esta clasificación y fueron modificando diversas dimensiones. Uno de los estudios más recientes engloba algunas terminologías ya conocidas y proponen otras, como ser [9], que presenta una lista de dimensiones para la calidad de los datos, y lleva a cabo encuestas anuales con el fin de medir el uso de las dimensiones de la calidad de los datos por parte de las organizaciones.

En base al relevamiento realizado, a continuación, se muestran los criterios que se tuvieron en cuenta como indicadores de validación de buena calidad para las métricas propuestas.

## **2 Trabajos relacionados sobre Métricas de calidad para datos abiertos**

Gobierno Abierto es un concepto que cada vez toma más énfasis en el ambiente de las organizaciones gubernamentales y en los ciudadanos a nivel mundial. De la mano de este contexto, también se hace un especial foco en los datos abiertos y la posibilidad de su utilización. En la Figura 1, se muestra la importancia del concepto de “datos abiertos” y cómo fue incrementando su tratamiento a través de los años, desde enero del 2004 a la actualidad. Para el desarrollo de dicho gráfico, se utiliza la herramienta Google Trends [10], en la que se pueden visualizar los números que reflejan el interés de búsqueda en relación con el valor máximo de un gráfico en un periodo determinados. Un valor de 100 indica la popularidad máxima de un término, mientras que 50 y 0 indican que un término es la mitad de popular en relación con el valor máximo o que no había suficientes datos del término, respectivamente.

Si bien existen muchos trabajos enfocados a Gobierno Abierto, y hasta algunos están orientados especialmente en datos abiertos, existen pocas investigaciones sobre el aspecto de calidad de los datasets disponibilizados en los portales gubernamentales. Algunas de las propuestas encontradas [11] [12] colaboran con el proceso de inculcar la cultura de gobierno abierto a las personas. Otros trabajos [13] [14], proponen un

prototipo de software para la evaluación de principios de datos abiertos que permitan, además, validar el cumplimiento de los datos abiertos.



**Fig. 1.** Interés de búsqueda del concepto “Datos Abiertos” en todo el mundo desde enero 2004 a la actualidad.

Otra propuesta [15] se orienta en la evaluación de portales de datos abiertos sobre criterios definidos por los autores, por ejemplo, como el planteamiento de metodologías e indicadores que miden la calidad de éstos para los sitios web de datos abiertos en algunos países. De este estudio se obtienen enfoques bien marcados: (a) Datos publicados: orientado a la calidad (disponibilidad, actualización, accesibilidad, visualización, formatos de publicación y completitud), uso (factores de demanda definida, número de visualizaciones, descarga, consumo de API y productos resultantes) y metadatos (identifican el factor de uso, completitud y Recuperabilidad); (b) Portales de datos abiertos: resaltando los aspectos de su estructura (identifican el factor de categorización), usabilidad (factor de búsqueda, Navegabilidad y consumo de datos) y mecanismos de comunicación. Por otro lado, otro trabajo [16] realiza foco en las técnicas de mediciones de aceptación, uso y grado de confianza de los usuarios. Varias investigaciones, utilizan los informes publicados por el sitio oficial del barómetro de datos abiertos reconocido internacionalmente [17], con el fin de medir el éxito y el impacto de las iniciativas de datos abiertos de los gobiernos. En aspectos de la evaluación de la usabilidad hay trabajos [18] que realizaron propuestas de buenas prácticas. Otros de los trabajos [19], [20], [21], [22], [23], [24] se enfocan en establecer criterios y clasificaciones de niveles de calidad de los datos abiertos.

### 3 Propuesta de Métricas de calidad de datos abiertos a través de HEVDA

Para este trabajo, se utiliza una herramienta desarrollada [14] llamada HEVDA (*HErramienta de Validación de Datos Abiertos*), que permite la validación de las distintas métricas sugeridas para un conjunto de datos abiertos en formatos CSV (valores separados por comas).

Para la elección del tipo de formato se tuvieron en cuenta estudios realizados por los autores [25], en los que se tomó como caso muestral el portal gubernamental Argentina Unida (Argentina Unida) con sus 973 conjuntos de datos hasta julio de 2020. Sus resultados concluyeron que el formato más utilizado es el tipo CSV con un 61,6% de uso, es por ello por lo que la herramienta HEVDA trabaja con el formato tipo CSV.

Si bien la herramienta HEVDA permite obtener un análisis automático, no modifica el dataset de origen, sino que brinda un estudio detallado que sirve como guía para la corrección de éste.

En la Figura 2, se muestra la pantalla inicial de la herramienta HEVDA (parte A, muestra el encabezado de ésta) con las métricas propuestas. Se puede observar la lista de métricas críticas (parte B) y no críticas (parte C) según el grado de importancia en calidad.



Fig. 2. Pantalla de inicio de la Herramienta HEVDA.

Las métricas sugeridas son parte de un análisis realizado sobre la Guía para la publicación de datos en formatos abiertos del Gobierno de la Ciudad Autónoma de Buenos Aires [26]. Las métricas que se analizan en este prototipo HEVDA son:

- a) **[MÉTRICA 1] Tratamiento de Números Decimales:** En la Figura 3, se muestra un ejemplo de la herramienta HEVDA, para la verificación de números decimales. Como se puede observar, en forma inicial existe una breve descripción de la métrica, y la cantidad de casos que se detectan para la validación del tipo de números decimales. Por ejemplo, “Hay 3 casos con números decimales mal cargados o separador decimal incorrecto con, (coma)”. Además, se disponibiliza un enlace “clic para ver detalle de registros”, el cual visualiza los casos afectados para este aspecto de calidad con una descripción mayor.



Fig. 3. Pantalla de Verificación de Números Decimales.

Posteriormente y de forma complementaria, se muestra un análisis estimativo para el cálculo de los tipos de datos de los campos del conjunto de datos analizados; esto se visualiza a través de una grilla que contiene: los nombres de los títulos de las columnas del dataset, y los tipos de datos detectados (según algoritmo interno propuesto en la herramienta HEVDA).

En la Figura 4, se muestra el detalle de casos que no cumplen con los criterios que se consideraron para el desarrollo de la herramienta (explicados anteriormente). Por ejemplo, se visualiza que el registro número 2, tiene un valor de “44.333” por lo que no cumple con el criterio del tipo cantidad de decimales menor o igual a 2, como así también, el valor “6.988” para el mismo registro. Para el caso del registro 3, se muestra el valor “11,43”, el cual no cumple con el criterio del tipo separador de decimal con “.” (punto).

Registros encontrados			
Nº Registro	Valor	Tipo	Columna Afectada
2	44.333	La cantidad decimales es mayor a 2	g
2	6.988	La cantidad decimales es mayor a 2	data
3	11,43	El separador decimal es incorrecto	g

Fig.4. Detalle de los casos detectados para la verificación de números decimales.

- b) **[MÉTRICA 2] Registros Duplicados:** para la verificación de registros duplicados (métrica de redundancia). Para ello se visualiza la cantidad de casos detectados con duplicación de registros, indicando la cantidad afectada sobre el total de registros del dataset.
- c) **[MÉTRICA 3] Datos Faltantes y/o Completos:** para identificar los datos que no se encuentran, identificando los registros de un dataset analizado.
- d) **[MÉTRICA 4] Caracteres inválidos:** para la detección de caracteres inválidos. Allí se visualiza la cantidad de columnas afectadas sobre el total de columnas del conjunto de datos, por ejemplo: “3 Columnas afectadas de un total de 9 del dataset”.
- e) **[MÉTRICA 5] Redundancia para el dominio de una columna:** para la medición de redundancia en el dominio de valores encontrados para una misma columna. Para ello se visualiza la cantidad de columnas detectadas de un dataset que tienen repetición en sus datos, es decir que tiene este mismo criterio de redundancia. Por ejemplo, se muestra la leyenda: “Se han encontrado: 7 columnas con campos duplicados de un total de 14 columnas”.
- f) **[MÉTRICA 6] Redundancia entre campos de una misma fila:** para la medición de redundancia entre los campos de una misma fila. Para ello se visualiza la cantidad de registros afectados que cumplen con valores repetidos sobre el total de registros del dataset. Por ejemplo, se muestra la leyenda: “78 Registros afectados de un total de 146 registros del dataset”.
- g) **[MÉTRICA 7] Detección de valores ID:** En la Figura 5, se muestra un ejemplo de la herramienta HEVDA para la estimación de campos con ID, detectándose los campos que contienen 'id', 'id\_' y/o '\_id', tanto en mayúsculas como en minúsculas, esta detección es importante para conocer el nivel de interoperabilidad o conexión con otra posible tabla de valores que identifiquen el significado de los ID, por ejemplo, “provincia\_ID”, podría estar identificado con una posible lista de Provincias asociadas a un código ID. En esta figura, es posible visualizar que se han encontrado 5 columnas que representan a ID (pais\_id; id; provincia\_id; categoria\_id; Certificadora\_id).



Fig.5. Métrica de Análisis de ID.

- h) [MÉTRICA 8] Campos Triviales:* Se identifican los nombres de las columnas que poseen en todos sus datos el mismo valor, por ejemplo, en la columna “País”, se presenta siempre el valor “Argentina”, por lo que es un dato que ocupa espacio en tamaño de almacenamiento (por ejemplo, si es un dataset con varios registros), con lo cual se podría agregar como nombre del archivo de un dataset, para eliminar el dato repetido en varias filas para la misma columna.

## 4 Criterios de Validación de las Métricas propuestas

Para este trabajo, por cada métrica propuesta, se la relacionó con una dimensión o criterio de calidad, que son el resultado de tomar en consideración, distintas fuentes:

- a) Norma ISO/IEC 25012 [27], que especifica un modelo general de calidad de datos que se encuentran definidos en un formato estructurado dentro de un sistema informático. Para este se presentan los criterios del modelo de calidad de datos definido por el estándar ISO/IEC 25012[27], de las 15 características que lo componen, se tomaron para este trabajo de investigación, los criterios de: ***Exactitud, Completitud, Consistencia y Precisión.***
- b) Estándar Universal de Calidad de Datos [28], son los criterios que debe contener un conjunto de datos para que puedan ser de calidad e interoperable y que son definidos por el estándar universal de la calidad de los datos de 2 capas. Para este aspecto se trabaja con el estándar universal en español extraído de la guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia [29], para la validación de las métricas propuestas, se consideran los criterios de:
  - a. Confidencialidad, siendo éstos: ***Precisión, Integridad, Consistencia y Completitud.***
  - b. Presentación, siendo éste el enfoque de ***estructura.***
- c) Dimensiones de la calidad de los datos (CDDQ) propuestas por Dan Myers en DQMatters [9]. Los criterios considerados para la comparativa de validación son: ***Completo, Exactitud, Consistencia, Integridad, Precisión y Representación.***
- d) Trabajos relevados enfocados a la calidad:
- e) Estudios realizados [25] en datasets abiertos de portales gubernamentales de la República Argentina.

### 4.1 Criterios de calidad utilizadas

En base a la sección anterior, a continuación, se detallan las dimensiones y criterios de calidad de datos tratados para el análisis, éstos son:



- a) **Precisión:** El criterio de precisión es el detalle de la medición que se utiliza para especificar un determinado dominio (conjunto de valores posibles) para un campo. Según [27], es el grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico. Básicamente, es el más alto nivel de desagregación posible, es decir, se evita hacer agregaciones ya que puede causar que se pierda lógica en la relación y análisis de datos asociados
- b) **Exactitud** [33]: que es el grado en el que los datos representan correctamente el verdadero valor del atributo deseado de un concepto o evento en un contexto de uso específico. Tiene dos enfoques, pero en este trabajo se tomará en cuenta la “Exactitud Sintáctica”, es decir, la cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado sintácticamente correcto [27]. Esto puede darse por un carácter inválido/especial.
- c) **Compleitud** [33]: se refiere a que, si el sistema de información contiene todos los datos de interés, y si los mismos cuentan con el alcance y profundidad que sea requerido. Esto conlleva a la existencia de dos factores de la completitud: cobertura y densidad:
  - a. La cobertura se refiere a la porción de datos de la realidad que se encuentran contenidos en el sistema de información;
  - b. La densidad se refiere a la cantidad de información contenida y faltante acerca de las entidades del sistema de información.
- d) **Integridad:** El criterio de integridad mide la calidad estructural o relacional de los conjuntos de datos, y se relaciona con los conceptos de validez, duplicación y coherencia [9]. Su enfoque es sobre el hecho de conocer si los datos son relevantes de forma que se puedan utilizar.
- e) **Consistencia:** El criterio de consistencia mide si los datos son equivalentes o no entre los sistemas o la ubicación de almacenamiento, también se relacionan con los conceptos de datos redundantes o distribuidos, coherencia de formato, coherencia lógica y coherencia temporal [9]. Según [27], es el grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto de uso específico. Puede ser analizada en datos que se refieran tanto a una como a varias entidades comparables.
- f) **Relación entre valores de campos:** El criterio de relación entre valores de campos se analiza desde la perspectiva de la capacidad de asociación de un valor con otro.
- g) **Estructurales/Representación:** El criterio de presentación o estructura, permite analizar los aspectos característicos de la estructura del conjunto de datos, básicamente orientado al archivo abierto. Según [9], la representación mide la facilidad de comprensión de los datos, la coherencia de la presentación, la elección de medios adecuados y la facilidad de leer e interpretar los datos.

- h) **Redundancia-Unicidad** [33]: Este aspecto indica el nivel de duplicación de los datos. Por ejemplo, cuando se trata de un aspecto que se encuentra representado en más de una vez en los datos analizados. Este conjunto de datos, representan exactamente el mismo valor u objeto, básicamente son los registros repetidos o redundantes. Por lo que la duplicación es cuando la misma entidad aparece repetida de manera exacta. Este término de unicidad [34], mide el grado en que un dato está libre de redundancias en amplitud, profundidad y alcance:
- a. En amplitud: las propiedades y clases representadas
  - b. En alcance: una base de conocimiento en donde múltiples ejemplares representan el mismo objeto
  - c. En profundidad: donde múltiples valores de una propiedad son únicos.

#### 4.2 Análisis de los criterios aplicados a las métricas sugeridas

En la sección 4 inicial, se presentaron los criterios de validación seleccionados para las métricas sugeridas en el presente trabajo. Como resultado de la sección anterior, se muestra la Tabla 1 con el análisis realizado, identificando tanto las métricas como los aspectos generales que influyen en uno o más criterios de calidad de datos. En las columnas se muestran los distintos criterios de validación de las métricas propuestas. Se indica con un punto si influye la métrica en el criterio.

**Tabla 1. Resumen del análisis de cotejo entre los criterios (Cr.) de calidad y las métricas propuestas (Mét.).**

Nro. Métrica	Precisión	Exactitud	Complejidad	Integridad	Consistencia	Relación	Estructura	Redundancia	Tot. Cr. x Mét.
1	•				•		•		3
2		•		•	•	•		•	5
3			•						1
4	•	•		•	•				4
5	•	•		•	•	•		•	6
6	•	•		•	•	•		•	6
7	•			•	•				3
8	•	•		•	•	•		•	6
<b>Tot. Mét. x Cr.</b>	6	5	1	6	7	4	1	4	
<b>Prom. Mét. x Cr.</b>	75%	62,50%	12,50%	75%	87,50%	50%	12,50%	50%	

En la Tabla 1, se puede observar que el criterio más representativo es el aspecto de Consistencia con un 87,50% (es decir, de las 8 métricas, se cumple en 7), que es el grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto de uso específico [25]. En lo que respecta a la Integridad y Precisión representan un 75%, siendo que la integridad enfoca la calidad estructural de los datasets, y se relaciona con la validez, duplicación y coherencia de éstos [45]. Como criterios menos relevantes, se visualizan los aspectos Estructurales y de Completitud con el 12,50%. Esto muestra cómo se cubrieron todos los criterios por medio de las métricas propuestas para este trabajo.

En la Figura 6 se muestra el promedio de la cantidad de las métricas propuestas que aplican o se relacionan con cada criterio de calidad.

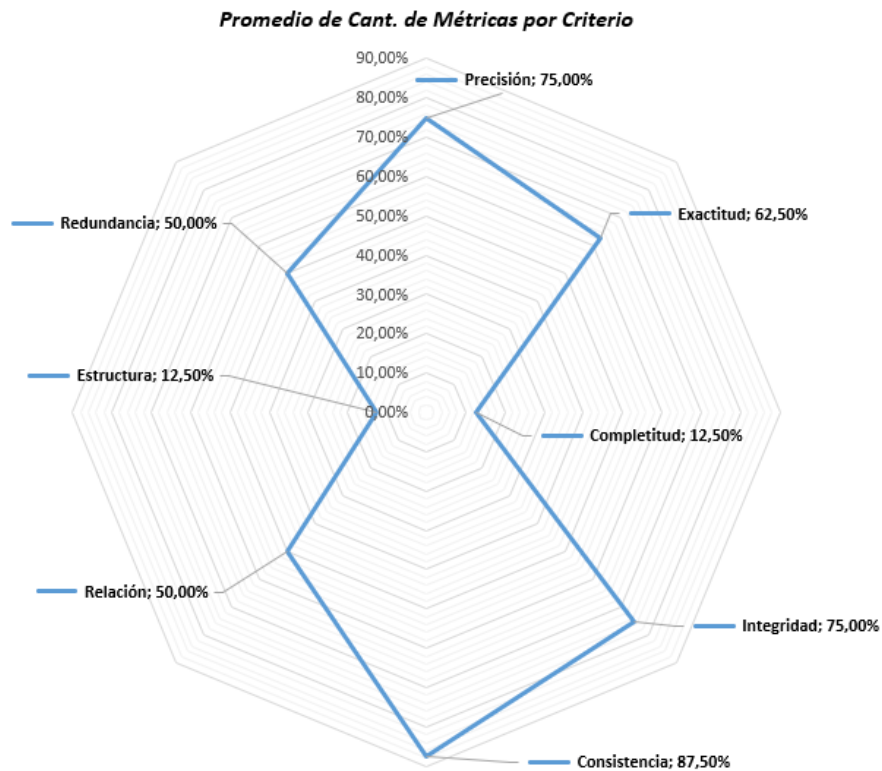


Fig. 6. Promedio de la cantidad de Métricas por Criterio de Calidad de datos.

Para la Figura 6, se visualiza que los criterios estandarizados de calidad que abarcan más tratamiento en el estudio de métricas propuestas son: la Consistencia, la Integridad, y la Precisión en calidad de datos.

## 5 Conclusiones

El papel de la tecnología es fundamental para concientizar y fomentar la utilización de los datos públicos gubernamentales en formatos abiertos, para lograr una mejor transparencia en el contexto de Gobierno Abierto y una mejor comunicación entre los ciudadanos y el Estado Nacional. Por tanto, es necesario considerar la importancia de la calidad de los datos abiertos de acuerdo con lo mencionado en este trabajo. Para lograr dicho objetivo, es vital enfocarse en métricas que ayuden a medir la calidad de los datos públicos abiertos disponibilizados como los conjuntos de datos en portales gubernamentales. Verificar un estudio con métricas permite conocer que tipo de “estado de salud de datos” existe en cuestiones de calidad.

Como resultado, y aporte de esta investigación, se presentaron las métricas propuestas para el análisis de datasets gubernamentales, las cuales fueron elaboradas según las buenas prácticas de la Guía para la publicación de datos en formatos abiertos del Gobierno de la Ciudad Autónoma de Buenos Aires [26], éstas fueron acompañadas e implementadas en un prototipo de desarrollo propio, aplicación HEVDA [14], el cual permite el estudio de un dataset y verificar que estado situación presenta para cada una de las métricas planteadas. Por otra parte, se mostró la validación de éstas métricas, a través de criterios estándares y dimensionales de calidad que respaldan el enfoque de la propuesta basadas en distintas fuentes: como Norma ISO/IEC 25012 [27], Estándar Universal de Calidad de Datos [28], Dimensiones de la calidad de los datos (CDDQ) propuestas por Dan Myers en DQMatters [9], trabajos relevados enfocados a la calidad, como así también en estudios realizados de los autores [25] en datasets abiertos de portales gubernamentales de la República Argentina.

Como futuras líneas de investigación se continuará estudiando el alcance e implementación de más métricas de calidad, así como su desarrollo en la herramienta HEVDA y así analizar y detectar más problemas en conjuntos de datos y mejorar tanto la calidad de los contenidos como la interoperabilidad.

## Referencias

1. Red Gealc – Red de Gobierno electrónico de América Latina y el Caribe, “Innovación Pública y Gobierno Abierto”, Fecha de publicación: marzo 2018, Disponible en: <http://www2.redgealc.org/contenido-general/noticias/innovacion-publica-y-gobierno-abierto/>
2. Olaya, Y. E. L. (2018), “Estudio sobre minería y visualización de datos abiertos del gobierno de Colombia”, Working papers, Maestría en Ingeniería de Sistemas, 2(2).
3. W3C (2009), “Publishing Open Government Data”, W3C Working Draft 8 September 2009, Disponible en: <https://www.w3.org/TR/gov-data/>
4. Cadena-Vela, S. (2019), “Marco de referencia para la publicación de datos abiertos comprensibles basado en estándares de calidad”, Tesis doctoral en Ciencias Informáticas, Universidad de Alicante.
5. Piñeiro, F. J. G. (2008), “Aspectos básicos de la calidad y de la gestión por procesos”, Lurralde: Investigación y espacio, (31), 277-289.

6. European Data Portal (2019), “Portal Europeo de Datos”, Disponible en: <https://www.europeandataportal.eu/es>
7. European Data Portal (2019), “Open Data Maturity Report”, Disponible en: [https://www.europeandataportal.eu/sites/default/files/open\\_data\\_maturity\\_report\\_2019.pdf](https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf)
8. The MIT Total Data Quality Management Program (2002), “MIT TDQM Program Highlight”, Disponible en: <http://web.mit.edu/tdqm/www/index.shtml>
9. Conformed Dimensions of Data Quality (2018), “Annual Survey about Use of Dimensions of Data Quality”, Disponible en: [http://dimensionsofdataquality.com/dims\\_survey](http://dimensionsofdataquality.com/dims_survey).
10. Google Trends, “Descubre qué está buscando el mundo”, Disponible en: <https://trends.google.es/trends/?geo=AR>
11. Rodríguez, J. A. M. (2019), “Valoración de factores de uso de los datos abiertos de gobierno”, Instituto de Ciencias de Gobierno y Desarrollo Estratégico (Doctoral Dissertation, Benemérita Universidad Autónoma De Puebla)
12. Vicente-Paños, A., & Jordán-Alfonso, A. (2017), “Acceso A La Información Pública Y Su Reutilización En Las Comunidades Autónomas: Evaluación De La Reutilización De Datos Abiertos”, *El profesional de la información*, 26(3).
13. Ariza Amado, D. F., & Rojas Clavijo, J. A. (2019), “Prototipo de Software para la evaluación de principios de datos abiertos”. Universidad Católica De Colombia, Facultad De Ingeniería, Programa De Ingeniería De Sistema, Trabajo De Investigación Tecnológica, Bogotá D.C., Colombia.
14. Martínez, R., & Parkinson, C. (2021), “Validación de la calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness”, Congreso Nacional de Ingeniería Informática y Sistemas de la Información, Universidad Nacional de Tecnología (UTN), Facultad Regional San Francisco, 2020.
15. Melo, C. A. H., & Sanabria, J. S. G. (2020), “Proposal for the Evaluation of Open Data Portals”, *Facultad de Ingeniería*, 29(54), 1-20.
16. Zainal, N. Z., Hussin, H., & Nazri, M. N. M. (2019), “Acceptance, Quality and Trust Factors–Conceptual Model for Open Government Data Potential Use”, *International Journal on Perceptive and Cognitive Computing*.
17. Open Data Barometer – World Wide Web Foundation, “The Open Data Barometer”, Disponible es: [https://opendatabarometer.org/?\\_year=2017&indicator=ODB](https://opendatabarometer.org/?_year=2017&indicator=ODB)
18. Máchová, R., Hub, M., & Lnenicka, M. (2018), “Usability evaluation of open data portals”, *Aslib Journal of Information Management*.
19. Oviedo Blanco, E. (2016), “Modelo de madurez para portales de datos abiertos e incorporación a la norma técnica nacional de Costa Rica”, Repositorio Institucional de la Universidad de Alicante, Tesis Doctoral.
20. Beltrán, L., Estefan, N., & Mahecha Moyano, J. F. (2017), “Prototipo de software para la evaluación de la calidad de datos abiertos”, Tesis de grado, Repositorio Institucional de la Universidad Católica de Colombia
21. Ibanez Gonzalez, L., Millard, I., Glaser, H., & Simperl, E. (2019), “An assessment of adoption and quality of linked data in European open government data”.
22. Rodríguez Rojas, L. A. (2017), “Metamodelo para integración de datos abiertos aplicado a inteligencia de negocios”, Tesis de Doctoral, Repositorio Institucional de la Universidad de Oviedo
23. Arizo, I. (2016), “Métricas basadas en datos”, Tesis de Maestría en Gestión de la Información, Universitat Politècnica de València.
24. Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2018), “Indicadores de calidad de datos abiertos: el caso del portal de datos abiertos de Barcelona”, *El profesional de la información (EPI)*.
25. Martínez, R., Rodríguez, R., & Vera, P. (2020, December). Analysis of datasets and catalogs in government open portals of the Argentine Republic. In 2020 IEEE Congreso Bienal de Argentina (ARGENCON) (pp. 1-8). IEEE.

26. Datos Gobierno de la Ciudad de Buenos Aires, “Estrategia de Apertura de Datos - Guía para la publicación de datos en formatos abiertos del Gobierno de la Ciudad Autónoma de Buenos Aires”, Disponible en: <https://datosgcba.github.io/guia-datos/guia-abiertos/>
27. ISO 25012 (2008), “Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos”, Disponible en: <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25012:ed-1:v1:en>
28. Cai, L., & Zhu, Y. (2015), “The challenges of data quality and data quality assessment in the big data era”, Data science journal, 14.
29. Guía de Estándares (2020), “Calidad e Interoperabilidad de los datos abiertos del Gobierno de Colombia”, Disponible en: [https://herramientas.datos.gov.co/sites/default/files/A\\_guia\\_de\\_estandares\\_final\\_0.pdf](https://herramientas.datos.gov.co/sites/default/files/A_guia_de_estandares_final_0.pdf)