Tipos de Métricas de calidad para validar datasets gubernamentales Argentinos

» **Roxana Martínez**
UIF, Financial Information Unit, Argentine.
Centro de Altos Estudios en Tecnología Informática (CAETI), Facultad de Tecnología Informática.
Universidad Abierta Interamericana (UAI), Ciudad Autónoma de Buenos Aires, Argentina.

Resumen

Hoy en día, los datos son un recurso esencial para toda actividad de la administración pública, y es necesario que estén actualizados, claros y completos. Si bien varios trabajos relacionados con este tema se centran en la apertura de los datos públicos en los portales de las instituciones gubernamentales, existen pocas investigaciones centradas en el contenido de los conjuntos de datos disponibilizados. El énfasis en el estudio de la calidad de los datos permite que las tecnologías mejoren su uso, manipulación y distribución a los ciudadanos. Lo que a su vez permite, una mayor participación ciudadana y transparencia de los entes estatales. Este trabajo se centra en la implementación de varias métricas estándares, que sostienen principios de criterios de calidad para ayudar en el correcto procesamiento e interoperabilidad de los datos. La utilización de métricas de calidad favorece al encuadre de indicadores que permitan obtener un dato más limpio para ayudar en el análisis final a través de análisis cuantitativo.

PALABRAS CLAVES: MÉTRICAS DE CALIDAD; DATOS ABIERTOS; GOBIERNO ABIERTO

Types of quality metrics to validate Argentine government datasets

Abstract

Today, data is an essential resource for all public administration activities, and it is necessary that it be up-to-date, clear and complete. Although several works related to this topic focus on the opening of public data in the portals of government institutions, there is little research focused on the content of the datasets made available. The emphasis on the study of data quality allows technologies to improve their use, handling and distribution to citizens. Which in turn allows greater citizen participation and transparency of state entities. This work focuses on the implementation of several standard metrics, which support quality criteria principles to help in the correct processing and interoperability of data. The use of quality metrics favors the framing

Revista Abierta de Informática Aplicada/vol 6 Nº2 (2022): 40-53 *Recibido: Junio 2022 / Aceptado: Noviembre 2022 / Publicado: Diciembre 2022*
ISSN 2591-5320

of indicators that allow obtaining cleaner data to help in the final analysis through quantitative analysis.

KEYWORDS: QUALITY METRICS; OPEN DATA; OPEN GOVERNMENT

1. Introducción

El nuevo paradigma de gobierno abierto permite llevar a cabo iniciativas por parte del Estado nacional que permitan un adecuado acceso a la información pública y promover instancias de co-creación de políticas públicas. Este enfoque desea mantener disponibles datos comunes compartidos, lo que requiere una gestión adecuada de las plataformas informáticas asociadas con éstos. Los ejes fundamentales para este contexto son:

a) ***Rendición de Cuentas*** en materia de políticas y desempeño de servicios públicos;

b) ***Transparencia***: lograr que los ciudadanos comprendan el funcionamiento de su gobierno;

c) ***Participación, colaboración y co-creación***: se enfoca en la participación activa de los ciudadanos en base a las políticas públicas;

d) ***Apertura de datos***: mantener disponible varias fuentes de conjuntos de datos públicos en formatos abiertos de distintas categorías, por ejemplo, educación, salud, economía entre otras;

e) Innovación: a través de las TICs (Tecnologías de la Información y la Comunicación) se podrán utilizar para la colaboración abierta en temas relacionados con la resolución de diversos problemas o bien para la resolución de problemas cívicos a través de sitios web públicos, con un enfoque que se centra en la cooperación y participación de los ciudadanos y las instituciones del Estado Nacional.

El gobierno abierto como concepto se hizo popular hace más de una década, especialmente con el discurso de Obama y posteriormente con la intervención de Open Government Partnership (OGP) [1]. Este contexto tiene como objetivo desafiar a los ciudadanos de una manera diferente, involucrándolos en los procesos de toma de decisiones y brindándoles nuevas herramientas para desempeñar un papel más significativo en la democracia [2].

En las siguientes secciones se analizarán los enfoques de innovación pública, la importancia del acceso a la información, la calidad del contenido de los datos públicos disponibles en los diversos sitios web gubernamentales y el contexto de la problemática actual.

1.1. Innovación Pública

El concepto de innovación pública se encuentra en pleno auge, tanto a nivel nacional como a nivel internacional. Este enfoque apunta al trabajo en lo que respecta a la inclusión y a la igualdad para mejorar la transparencia en la administración pública, de forma tal que permite garantizar la conectividad a través de las distintas plataformas tecnológicas, y, por otro lado, trabajar en el tratamiento y cumplimiento del acceso a la información pública para todos los ciudadanos.

La Secretaría General Iberoamericana (SEGIB) [3] "coincidiera que la innovación pública se ha transformado en los últimos años en un proceso necesario para modernizar las administraciones, universidades, parlamentos, organismos internacionales y otros tantos estamentos. Esto supone poner al sector público como un agente capaz de liderar el desarrollo sostenible, de una forma más sincronizada con otros sectores" [4].

La innovación pública implica trabajar con los ciudadanos para crear soluciones y procesos, es decir, identificar sus necesidades y escucharlos. "La necesidad de nuevas capacidades potencialmente obtenibles desde el exterior de las estructuras estatales clásicas por medio de la cocreación ciudadana, conecta estos nuevos problemas que surgen en la escena contemporánea con el concepto de innovación pública —cuyos tres pilares son la innovación abierta, la cocreación y la inteligencia colectiva—, y los laboratorios de innovación son el espacio idóneo para conjugar estas tres características como espacios-frontera" [5]. Estos espacios-fronteras permiten la comunicación entre diferentes actores con diferentes saberes, consenso y cooperación ciudadana.

1.2. Acceso a la Información Pública

Debido a toda esta información disponible, los gobiernos han aprobado diversas leyes que brindan el marco legal para la privacidad de los datos disponibles y reconocen que el acceso a esta información es un derecho de los ciudadanos, otorgándole el derecho de permitir el control, seguimiento y toma de decisiones colaborativa.

De acuerdo con el Decreto 434/2016 [6], la estrategia nacional se estructura entorno a cinco ejes: a) tecnologías digitales y planificación pública; b) gestión integrada de recursos humanos; c) gestión de la aplicación y participación pública; d) gobierno abierto e innovación social; y e) Estrategia Nacional Digital.

De los cinco ejes mencionados, las TICs se utilizan de forma prioritaria en dos, por ejemplo:

1. ***Iniciativa de Tecnología y Gobernanza Digital***: planes para incluir la infraestructura tecnológica y las redes necesarias para avanzar en nuevos modelos de gobernanza basados en tecnologías de big data;

2. ***Gobierno abierto e innovación pública***: no solo complementa la base anterior, sino que permite evaluar y controlar la política pública del país. En otras palabras, procesar grandes cantidades de información es un recurso valioso y necesario.

Por tanto, partiendo de esta premisa, es muy necesario gestionar esta estructura mediante el uso e integración de nuevas herramientas de análisis de datos.

Es importante comprender que este acceso, "permite una mayor transparencia, ofrece servicios públicos más eficientes, y alienta un mayor uso público y comercial, en 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" [7]. Por otra parte, a transparencia dentro de lo que es el contexto de Gobierno Abierto, consiste en asegurar el derecho de todos los ciudadanos, que es el libre acceso a la información pública gubernamental. De esta manera, un gobierno puede mostrar de manera sencilla y clara, la gestión llevada a

Revista Abierta de Informática Aplicada/vol 6 Nº2 (2022): 40-53
ISSN 2591-5320

cabo, y así promover la gestión activa. Este derecho a la información pública permite conocer y utilizar datos que producen o tienen los tres poderes del Estado, la Ley que ampara este acceso es la Ley 27.275 [8].

1.3. Calidad de los datos abiertos gubernamentales

Para este nuevo paradigma, las tecnologías de la información son uno de los elementos claves para llevar a cabo el desarrollo, la integración de nuevas tecnologías informáticas y de telecomunicaciones dentro del Estado nacional, para optimizar la gestión del sector público.

Distintos sitios web gubernamentales ponen a disposición de los ciudadanos, diversos conjuntos de datos públicos en formatos abiertos, llamados datasets [9]. Éstos corresponden a diversas categorías enfocadas al interés nacional, por ejemplo: datos de salud, educación, seguridad, gobierno y sector público, justicia, seguridad y legales, ciencia y tecnología, energía, medio ambiente, transporte entre otras.

Esta apertura de datos promueve la colaboración entre el estado y la ciudadanía, lo que permite la co-creación de soluciones innovadoras que benefician a la sociedad, como así también, visualizan aspectos de transparencia gubernamental y permitir nuevas oportunidades de desarrollo ciudadano.

Además de los conjuntos de datos públicos abiertos, existen servicios de datos en varios sitios gubernamentales, Estos servicios proveen el acceso a datos a través de APIs de uso libre que pueden ser integradas en diversas aplicaciones de software [10]. Conjuntamente, se brindan herramientas que facilitan la apertura de datos, algunas de estas herramientas son guías de buenas prácticas para la apertura de datos en organismos de la administración pública nacional [11] [12], o bien guía para la publicación de datos en formatos abiertos [13], o bien estándares para que las diversas organizaciones puedan documentar sus activos de datos de forma práctica y homogénea para la para toda la administración pública nacional, es decir utilizando metadatos para datos abiertos [14] [15].

Aunque los portales de datos abiertos proporcionan muchos conjuntos de datos, tanto nacionales como internacionales, ha habido una serie de casos en los que carecían de datos de calidad, es decir, tenían un formato incorrecto, tenían errores ortográficos o encontraban campos incompletos. Por su parte, algunas organizaciones apoyan aspectos y/o principios de calidad de datos.

Es necesario que varias entidades gubernamentales, recopilen, compartan y utilicen datos bien organizados y administrados para impulsar políticas y prácticas que aborden de manera efectiva los desafíos sociales, económicos y ambientales más apremiantes. Aunque existen muchos organismos [16][17], [18] con objetivos similares, es decir, proporcionar principios similares para identificar datos gubernamentales como datos abiertos, la mayoría respaldan la idea de que los datos gubernamentales deben estar disponibles para todos los ciudadanos, ya que son financiados por estos fondos públicos y, por lo tanto, siempre deben estar disponibles en un formato abierto que cumpla con los estándares técnicos y permita su reutilización y distribución. El sistema de modernización en Argentina "propone una gestión innovadora con una plataforma tecnológica de datos abiertos y un acceso a la información integrada, la cual es aplicada a toda la administración estatal. Para llevar a cabo dicho proyecto, es de vital importancia,

la concientización de las personas frente a esta nueva visión de acceso a la información pública, brindando una propuesta en tiempo real, en forma clara y precisa" [19]. Para ello, garantizar la calidad de los datos disponibilizados no es sencillo [20], por lo que estudiar los procesos de recolección de estos, como así también de administración y disponibilización, puede ayudar a un buen estudio de resultados.

1.4. Contexto del Problema

Algunos gobiernos han implementado diferentes métodos de gestión para las situaciones antes mencionadas, sin embargo, aplicar estos métodos y utilizar la digitalización en los diversos procesos que existen en la actualidad, no es una solución completa.

Si bien este paradigma se ha establecido en varios países del mundo, existe la necesidad de incluir una mayor concientización y disponibilidad de conjuntos de datos en varias organizaciones, y que, además, diversas entidades aún no se han involucrado en este tema. Uno de los principales problemas que se encuentran en estos portales de datos abiertos es la inconsistencia de algunos conjuntos de datos, lo cual se debe a que no todos los datasets cumplen con un estándar determinado, sino que las organizaciones publican de acuerdo con las pautas recomendadas y/o sus propios estándares predefinidos, es decir, las organizaciones pueden tener estándares diferentes. básicamente no estándares internacionales obligatorios consensuados con otros organismos. Este inconveniente lleva a que varios datos tengan un bajo nivel de calidad, incrementando los problemas de reutilización y análisis de éstos. "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" [21].

2. Trabajos relacionados

En esta sección se analizan algunos de los trabajos relacionados con respecto a la calidad de los portales de datos abiertos.

Varios trabajos [22] [23], [24], [25], [26] [27] mencionan la necesidad de contar con técnicas y estrategias para verificar la calidad de los datos que se brindan a la comunidad, con el fin de obtener un análisis que ofrezca un valor agregado a la sociedad. Además, "diversas investigaciones indican que es importante en enfocarse en la apertura de datos de calidad, prestando atención a aspectos como infraestructura, interoperabilidad, herramientas de publicación y licencias de uso adecuado. Otro punto es la presentación de los datos en formatos amigables, para que los datos sean accesibles, disponibles en línea, preferiblemente de forma gratuita y sin necesidad de registrarse, sin barreras administrativas, tecnológicas o burocráticas que restrinjan su acceso" [28]. Otros trabajos [29], [30] [31] [32], [33], [34] definen mecanismos adecuados para evaluar la madurez de un portal de datos abiertos a través de métricas de calidad, como, por ejemplo: la trazabilidad, la completitud y la conformidad, o bien orientan la evaluación de la calidad de los portales abiertos en el análisis del modelo de cinco estrellas de Berners-Lee [35] y otros factores que ayudan a evaluar la calidad para una adecuada reutilización.

3. Propuesta de métricas de calidad

Este trabajo es una ampliación de las métricas propuestas que se basan de trabajos anteriores [28] [36] [37] [38].

Varios conjuntos de datos se encuentran disponibles en los portales de los gobiernos nacionales e internacionales. Muchas veces estos datos no son de buena calidad, es decir, no están en formato abierto, existen caracteres inválidos o bien campos incompletos.

La propuesta de este trabajo es generar métricas que permitan el análisis de estos aspectos, ya que los datos son un recurso esencial para cualquier actividad de la administración pública y, por lo tanto, deben actualizarse y complementarse. Si bien está claro que cada vez más gobiernos están adoptando este nuevo paradigma de conceptos de datos abiertos, todavía queda un largo camino por recorrer. La importancia de la calidad de los datos favorece a una mejor utilización, manipulación y distribución a los ciudadanos para varios fines, permitiendo a su vez, una mayor participación ciudadana y transparencia en organismos estatales.

Para analizar la calidad de datos es necesario cuantificar determinadas características en el conjunto de datos analizados. Para el estudio de la calidad de datos, este trabajo se orienta en algunos aspectos más relevantes en cuestiones de calidad de datos sobre criterios estándares de la Norma ISO/IEC 25012 [39], estándares universales de calidad de datos [40], dimensiones de la calidad de los datos [41], trabajos relevados y estudios realizados [42] en esta temática, como los enfocados en la calidad del dato (Normas y estándares) y guías elaboradas por el Gobierno de la Ciudad de Buenos Aires [43], que plantean una serie de criterios para tener en cuenta en la apertura y tratamiento del contenido de los datasets públicos.

A modo de propuesta de métricas de calidad se identifican grupos de éstas. Cada una de las clasificaciones enfocan en aspectos determinantes en el contenido de los datasets, a continuación, se explica cada categoría:

a) ***Métricas críticas estándares:***
   Contienen aquellas métricas que permiten detectar problemas de datos de una índole prioritaria para un correcto análisis de resultados con datasets, como ser: cuestiones de redundancia, contenido faltante en registros o bien datos erróneos. Es decir, es necesario tener en cuenta estos aspectos, ya que su presencia no favorece a un correcto estudio de los datos disponibilizados.

b) ***Métricas no críticas estándares:***
   Contienen aquellas métricas que pudieran representar problemas de contenido en el dataset. Su detección está enfocada a posibles estimaciones de casos de errores y datos triviales, como así también, descubrimientos de datos redundantes combinados (entre campos y/o registros del dataset) que podrían conducir a inconvenientes en el análisis de un conjunto de datos.

c) ***Métricas para datos abiertos geoespaciales:***
   Contienen aquellas métricas que se enfocan en los datos georeferenciados, como ser: datos cartográficos enfocados en regiones del país, mapas de clima, radio geográfico de

una determinada cobertura en aspectos de servicios públicos, entre otras. estas métricas pretenden analizar la estructura de este tipo de datos disponibilizados en los sitios de datos abiertos.

*d) **Métricas de catálogos de datos abiertos:***
Contienen aquellas métricas que se utilizan para analizar la lista de activos de datos y sus metadatos, ya que los catálogos de los datasets disponibles en los portales gubernamentales, describen el contenido de los conjuntos de datos y cómo pueden ser utilizados por otros organismos, es por esto por lo que, es importante definir métricas para la validación de la estructura de la metadata de éstos.

En la Figura 1, se indican los cuatro grupos de métricas para la medición del contenido de un conjunto de datos público abierto, siendo éstas: a) Métricas críticas estándares; b) Métricas no críticas estándares; c) Métricas para datos abiertos geoespaciales; y d) Métricas de catálogos de datos abiertos.

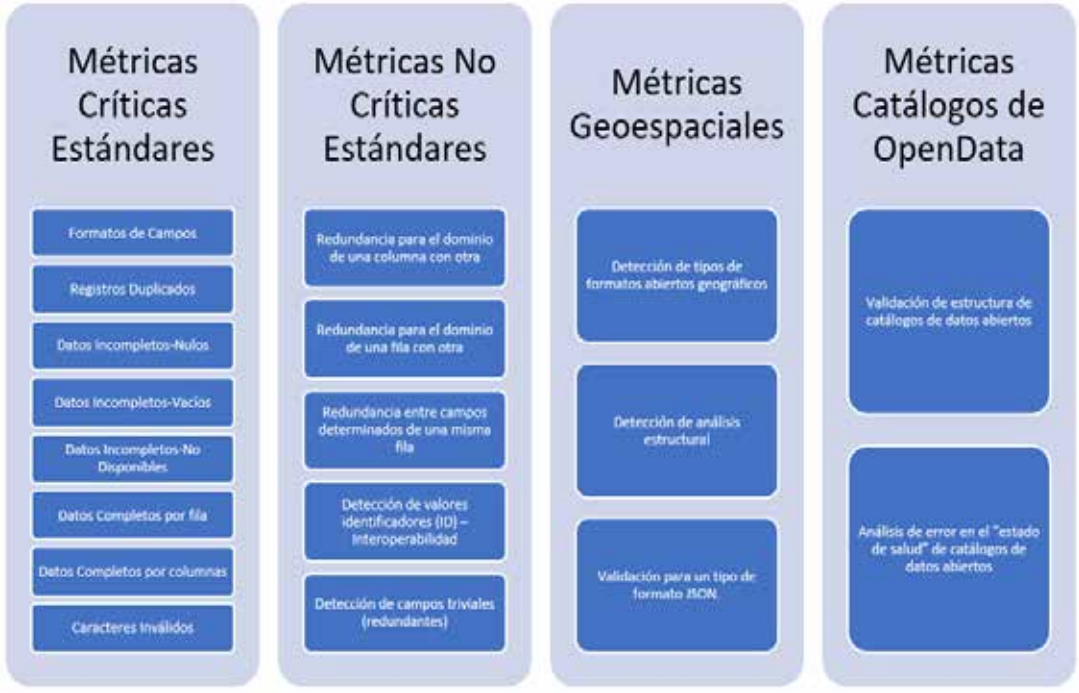


Figura 1. *Clasificación de Métricas de calidad propuestas para el contenido de datos abiertos públicos gubernamentales.*

Dichas métricas son implementadas en una herramienta desarrollada, la cual es nombrada "HEVDA, Herramienta de Validación De Datos Abiertos". Mediante esta aplicación, en trabajos anteriores [28] [37], se realizaron diversas pruebas a datasets seleccionados de algunos portales gubernamentales de datos abiertos más relevantes de Argentina, los cuales formaron parte de la recolección de datos definidos como muestra. Dicho estudio, permitió orientar la propuesta de métricas identificadas, ya que se realizó en base al análisis detallado de los resultados extraídos. Cabe mencionar que la herramienta desarrollada que permite la validación de las distintas métricas sugeridas para un conjunto de datos abiertos en formatos CSV (valores separados por comas) permite obtener un análisis automático, por lo que no modifica el dataset de origen

analizado, sino que brinda un estudio detallado ("estado de salud"), que sirve como guía práctica orientativa para la corrección de este.

La funcionalidad general del software HEVDA consiste en seleccionar un archivo dataset del tipo de formato CSV y efectuar la validación de las métricas propuestas para las dos primeras clasificaciones, actualmente se está trabajando en un prototipo actualizado para las dos últimas clasificaciones de métricas propuestas.

4. Validación de la medición de calidad en datasets

La validación de las métricas propuestas se basa en varios aspectos definidos. Esta definición surge de la investigación de ISO/IEC 25012, del Estándar común de calidad de datos, y sobre diversos trabajos de investigación relevados.

4.1. Consideraciones

Por cada métrica propuesta, se la relacionó con una dimensión o criterio de calidad, que son el resultado de tomar en consideración, distintas fuentes [28]:

a) Norma ISO/IEC 25012 [39]***:*** Indica un modelo general de calidad de datos definido en un formato estructurado en un sistema informático. Para este se presentan los criterios del modelo de calidad de datos definido por el estándar ISO/IEC 25012 [39], de las 15 características que lo componen, se tomaron para este trabajo, los criterios de: Exactitud, Completitud, Consistencia y Precisión.

b) Estándar Universal de Calidad de Datos [40]: 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. Para la validación de las métricas propuestas, se consideran los criterios de:

- Confidencialidad, siendo éstos: Precisión, Integridad, Consistencia y Completitud;
- Presentación, siento éste el enfoque de estructura.

c) Dimensiones de la calidad de los datos (CDDQ): propuestas por Dan Myers en DQMatters [41]. Los criterios considerados para la comparativa de validación son: Completo, Exactitud, Consistencia, Integridad, Precisión y Representación.

d) Trabajos relevados [28] [36] [37] [38].

A continuación, se muestra la Tabla 1, en la que se presentan los criterios de comparación seleccionados y propuestos para las métricas sugeridas:

Tabla 1. Criterios de Validación y sus aspectos descriptivos.

Criterios de Validación	Aspectos descriptivos para tener en cuenta
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 [39], 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.
Exactitud	El criterio de exactitud 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 principales aspectos de los cuales se considerará 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 [39].
Completitud	El criterio de completitud es el grado en el que los datos asociados con una entidad tienen valores para todos los atributos esperados e instancias de entidades relacionadas en un contexto de uso específico [39].
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 [39]. Su enfoque es sobre el hecho de conocer si los datos son relevantes de forma que se puedan utilizar.
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 [41]. Según [39], 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.
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.
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 [41], 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.
Redundancia	Permite analizar si se tienen los mismos valores en formatos iguales o similares dentro de un archivo. 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.

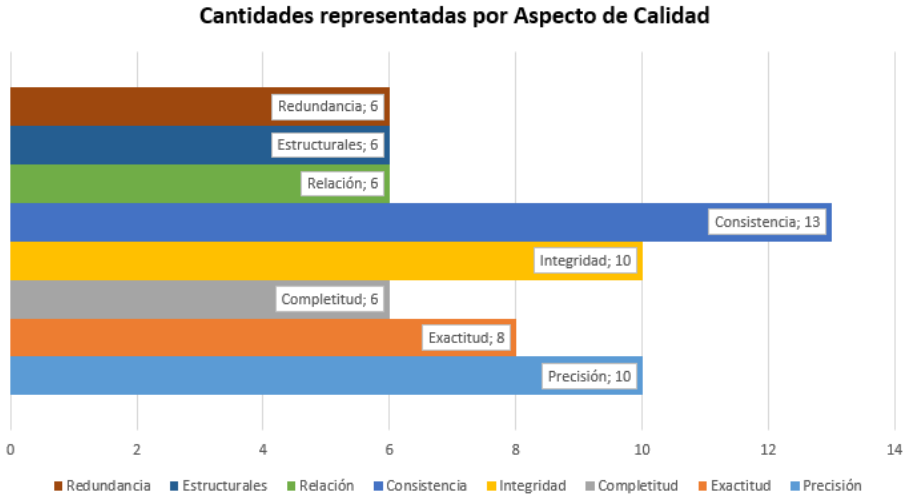
A continuación, se muestra la Tabla 2 con el resumen del análisis, identificando que todas las métricas como los aspectos generales influyen en uno o más criterios.

Tabla 2. Resumen de los criterios de validación y la implicancia en cada métrica propuesta.

Nro. Métrica	Tipo Métrica	Métricas Propuestas	Precisión	Exactitud	Completitud	Integridad	Consistencia	Relación	Estructurales	Redundancia
1	1	Formatos de Campos	X				X		X	
2		Registros Duplicados		X		X	X	X		X
3		Datos Incompletos – Nulos			X					
4		Datos Incompletos – Vacíos			X					
5		Datos Incompletos – No Disponibles			X					
6		Datos Completos por fila			X					
7		Datos Completos por columnas			X					
8		Caracteres Inválidos	X	X		X	X			
9	2	Redundancia para el dominio de una columna con otra	X	X		X	X	X		X
10		Redundancia para el dominio de una fila con otra	X	X		X	X	X		X
11		Redundancia entre campos determinados de una misma fila	X	X		X	X	X		X
12		Detección de valores identificadores (ID) – Interoperabilidad	X			X	X			
13		Detección de campos triviales (redundantes)	X	X		X	X	X		X
14	3	Detección de tipos de formatos abiertos geográficos	X				X		X	
15		Detección de análisis estructural				X	X		X	
16		Validación para un tipo de formato JSON		X		X	X		X	
17	4	Validación de estructura de catálogos de datos abiertos	X				X		X	
18		Análisis de error en el "estado de salud" de catálogos de datos abiertos	X	X	X	X	X	X	X	X

Se presentaron los criterios de validación seleccionados para las métricas sugeridas con el resumen del análisis, identificando que todas las métricas como los aspectos generales influyen en uno o más criterios.

Figura 2. Cantidad influyente de los criterios de calidad en las métricas propuestas.



Como resultado se muestra la Figura 2, en el que se visualizan los 3 criterios de validación con más influencia en las métricas sugeridas, siendo éstas: la consistencia, la integridad y la precisión en el contenido de los datasets.

5. Conclusiones

La idea de gobierno abierto no es un tema nuevo, pero toca varios conceptos que están surgiendo en todo el mundo en aspectos de administración pública, teniendo como temas centrales pilares a la: transparencia, innovación, participación ciudadana, rendición de cuentas y colaboración. El gobierno debe estar abierto a los ciudadanos, porque tienen derecho a todos los recursos necesarios en términos de responsabilidad, además de la oportunidad de poder participar en la formación e implementación de políticas nacionales como internacionales.

Este trabajo se centra en la detección de falencias en los contenidos de los conjuntos de datos públicos, que se encuentran en un formato abierto en sitios web gubernamentales. Por lo que, se brinda un relevamiento de la situación actual de los aspectos más relevantes en innovación pública dentro de este contexto, y se proponen un conjunto de métricas para analizar la calidad de los contenidos de los conjuntos de datos. Por otra parte, se analizan varios aspectos de estándares, buenas prácticas, guías de referencia, normas y dimensiones de calidad que son importantes para tener en cuenta a la hora de llevar a cabo la detección de falencias en una validación de contenido. Realizar a tiempo un análisis del contenido de los datos disponibles, antes que sean publicados en sitios web, permitirá evitar anomalías o bien inconsistencias en la comprensión del contenido brindado a los ciudadanos.

Disponer de diversas guías que orienten la mejora constante de calidad de los datos abiertos es fundamental, pero, además, es vital contar con herramientas que permitan una rápida validación para tener una mejor visualización en cuanto a falencias o falta de integridad en los conjuntos de datos con el fin de aplicar las mejoras correspondientes en estos. Tener a disposición datos públicos abiertos de calidad, permitirá a los ciudadanos, una mejor confianza en las fuentes de datos y seguimiento de procesos administrativos del Estado Nacional. Básicamente, definir diversas métricas que analicen distintas propiedades del contenido de los conjuntos de datos, tales como repeticiones, tipos de redundancias, validaciones de caracteres y otras, permitirá analizar los datos de una mejor manera, con el fin de evitar que conduzcan a estudios falsos sobre datos incorrectos o sucios.

Los resultados de los sondeos realizados en diversas investigaciones evidencian que es posible definir métricas de calidad de datos públicos gubernamentales en un formato abierto para efectuar un análisis cuantitativo a través de una herramienta amigable, sencilla e intuitiva para tratar aspectos de buenas prácticas.

Referencias:

» [1] Open Government Partnership (OGP). (2022). Committed to making governments more open, accountable, and responsive to citizens. Disponible en: https://www.opengovpartnership.org/

» [2] Gobierno de España – Aprendizaje Conectado. (2021). Fundamentos del Gobierno Abierto, una nueva experiencia de aprendizaje en línea. Disponible en: https://www.aprendizajeconectado.inap.info/2021/10/fundamentos-del-gobierno-abierto-una.html

» [3] Secretaría General Iberoamericana (SEGIB). (2022). Secretaría General Iberoamericana (SEGIB). Disponible en: https://www.segib.org/

» [4] Oliván Cortés, R. (2020). Instituciones que aprenden. HIP: un modelo de innovación pública para la era post-COVID. Resumen ejecutivo. Madrid: SEGIB. Retrieved July, 3, 2020.

» [5] Rodríguez, Exequiel. (2018). Laboratorios de innovación pública. Estado Abierto. Revista sobre el Estado, la administración y las políticas públicas. vol. 3. no 1.

» [6] InfoLeg Información Legislativa. (2016). Decreto 434/2016. Disponible en: http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/259082/norma.htm

» [7] W3C. (2009). Publishing Open Government Data. W3C Working Draft 8 September 2009. Disponible en: https://www.w3.org/TR/gov-data/

» [8] Infoleg - Información Legislativa. (2016). Derecho de Acceso a la Información Pública - Ley 27275. Sancionada por el Congreso el 14 de septiembre de 2016. Disponible en:

» http://servicios.infoleg.gob.ar/infolegInternet/anexos/265000-269999/265949/norma.htm

» [9] datos.gob.ar. (2022). Datos Argentina. Disponible en: https://datos.gob.ar/

» [10] datos.gob.ar. (2022). APIs. Disponible en: https://datos.gob.ar/apis

» [11] datos.gob.ar. (2022). Guía para la apertura de datos en organismos de la Administración Pública Nacional. Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia-apn/

» [12] datos.gob.ar. (2022). Guía para la identificación y uso de entidades interoperables. Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_interoperables/

» [13] datos.gob.ar. (2022). Guía para la publicación de datos en formatos abiertos. Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/

» [14] datos.gob.ar. (2022). Perfil de Aplicación Nacional de Metadatos para Datos Abiertos. Disponible en: https://datosgobar.github.io/paquete-apertura-datos/perfil-metadatos/

» [15] Argentina.gob.ar. (2022). Datos Abiertos. Disponible en: https://www.argentina.gob.ar/jefatura/innovacion-publica/servicios-y-pais-digital/gobierno-abierto/datos-abiertos

» [16] The International Open Data Charter. (2022). Open Data Charter. Disponible en: https://opendatacharter.net/

» [17] Open Government Partnership. (2022). Argentina: Miembro desde 2012 | Plan de Acción 4. Disponible en: https://www.opengovpartnership.org/es/members/argentina/

» [18] Open Government Data. (2022). The 8 Principles of Open Government Data. Disponible en: https://opengovdata.org/

» [19] Martinez, R., Rodríguez, R. A., & Vera, P. M. (2018). Análisis del diseño adaptativo en sitios web gubernamentales. In XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018).

» [20] Martínez, R., Rodríguez, R. A., Vera, P. et al. (2019). Análisis de técnicas de raspado de datos en la web aplicado al Portal del Estado Nacional Argentino. In XXV Congreso Argentino de Ciencias de la Computación (CACIC). (Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019).

» [21] Martínez, R., Pons, C. F., Rodríguez, R., & Vera, P. M. (2021). Validación de Métricas propuestas de Calidad para el estudio de datos abiertos en base a criterios estándares: Aplicación HEVDA. In XXII Simposio Argentino de Ingeniería de Software (ASSE 2021)-JAIIO 50 (Modalidad virtual).

» [22] Muente-Kunigami, A., & Serale, F. (2018). Los datos abiertos en América Latina y el Caribe. Banco Interamericano de Desarrollo (BID). Disponible en: https://gobiernoabiertogto.org.mx/documentos/

libros/48.pdf

» [23] 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. Disponible en: http://rua.ua.es/dspace/handle/10045/54316

» [24] Beltrán, L., Estefan, N., & Mahecha Moyano, J. F. (2017). Prototipo de software para la evaluación de la calidad de datos abiertos. Repositorio Institucional de la Universidad Católica de Colombia. Disponible en: Data Quality in OpenData.pdf (ucatolica.edu.co)

» [25] Ibanez Gonzalez, L., Millard, I., Glaser, H., & Simperl, E. (2019). An assessment of adoption and quality of linked data in European open government data.

» [26] 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. Disponible en: http://digibuo.uniovi.es/dspace/handle/10651/44552

» [27] Arizo, I. (2016). Métricas basadas en datos. Tesis de Maestría en Gestión de la Información. Universitat Politècnica de València.

» [28] Martínez, R. (2022). Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales (Doctoral dissertation, Universidad Nacional de La Plata).

» [29] 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), 27(2), 375-382.

» [30] Nayek, J. K. (2018). Evaluation of Open Data Government Sites: A Comparative Study. Library Philosophy & Practice. (e-journal). 1781. Disponible en: https://digitalcommons.unl.edu/libphilprac/1781

» [31] 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.

» [32] Cadena-Vela, S., Fuster-Guilló, A., & Mazón, J. N. (2019). Publicando datos abiertos considerando criterios de calidad.

» [33] Leonangeli, S., & Maenza, R. R. (2019), "Hacia un paradigma de calidad de Datos Abiertos adecuado al propósito de Gobierno Abierto", In XIII Simposio de Informática en el Estado (SIE 2019)-JAIIO 48 (Salta).

» [34] Schieferdecker. (Open) Data Quality. 2012 IEEE 36th Annual Computer Software and Applications Conference. Izmir. Turquía. 2012. págs. 83-84, doi: 10.1109 / COMPSAC.2012.120

» [35] 5 Open Data (2012). 5 Open Data. Disponible en: https://5stardata.info/en/

» [36] Martínez, R., Rodríguez, R. A., & Vera, P. M. (2021). Metrics proposal to measure the quality of governmental datasets. IEEE Latin America Transactions, 20(2), 301-308.

» [37] Martínez, R., Pons, C., Rodríguez, R., & Vera, P. (2021). Quality evaluation of government open data sets in Argentina using the HEVDA Validation Tool. Journal of Science and Research, 6(2).

» [38] Martínez, R. et al. (2021). Análisis de la Apertura de datos gubernamentales en los portales provinciales de la República Argentina: Aplicación HEVDA. Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI.)

» [39] 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

» [40] Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. Data science journal. 14.

» [41] Conformated Dimensions of Data Quality (2018). Annual Survey about Use of Dimensions of Data

Quality. Disponible en: http://dimensionsofdataquality.com/dims_survey

» [42] 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.

» [43] Secretaría de Modernización. Presidencia de la Nación. (2022). Paquete de Apertura de Datos de la República Argentina. Disponible en: Guía para la apertura de datos en gobiernos provinciales y locales - Paquete de Apertura de Datos de la República Argentina (datosgobar.github.io)