

Analysis of datasets and catalogs in government open portals of the Argentine Republic

Roxana Martínez

Centro de Altos Estudios en Tecnología
Informática (CAETI), Universidad
Abierta Interamericana (UAI).
Ciudad Autónoma de Buenos Aires,
Argentina
roxana.martinez@uai.edu.ar

Rocío Rodríguez

Centro de Altos Estudios en Tecnología
Informática (CAETI), Universidad
Abierta Interamericana (UAI).
Ciudad Autónoma de Buenos Aires,
Argentina
rocioandrea.rodriguez@uai.edu.ar

Pablo Vera

Centro de Altos Estudios en Tecnología
Informática (CAETI), Universidad
Abierta Interamericana (UAI).
Ciudad Autónoma de Buenos Aires,
Argentina
pablmartin.vera@uai.edu.ar

Abstract—This paper analyzes the quality of the data in the data sets presented in the open portal sites of the Argentine Republic, considering the portal of the Autonomous City of Buenos Aires (C.A.B.A) and the decentralized organizations of the country. The objective is to show results compared to the analysis of the data sets published in the open portals, with some important considerations when maintaining, managing and reusing systems, in order to generate added value to the community. Finally, the open data catalogs are validated, using a Python library (pydatajson), to verify if they comply with a standard catalog structure, proposed by the government of the Argentine Republic.

Keywords— Open data, Public data, Data catalogs, Open Government

I. INTRODUCCIÓN

La aplicación de las tecnologías de la información y la comunicación (TIC), junto con la mejora de varios procesos existentes hoy en día en el gobierno público, no presentan una solución en su totalidad. Esto es debido a que existen grandes déficits en algunas funciones que están determinadas por modelos administrativos antiguos, en el que muchos ministerios trabajaban con un paradigma orientado a una administración únicamente entre sectores departamentales, en lugar de tener una mirada enfocada en lo interdisciplinario y colaborativo junto con los ciudadanos. Esta premisa genera una caída en la legitimidad del sistema político en general y fomenta un contexto de desconfianza por parte de los ciudadanos. La problemática de este paradigma reside en la ausencia de información, eficiencia y transparencia, y, además, por la falta de respuestas públicas rápidas y eficientes. Debido a lo anteriormente explicado, el Estado Argentino enfrenta nuevos desafíos en el rediseño de innovadoras políticas públicas. “Tanto el gobierno nacional como los provinciales y municipales están admitiendo, al menos discursivamente, que el Estado debe abrirse a la sociedad, eliminar el secreto, hacer pública toda la información disponible, escuchar al ciudadano y promover su rol en la formulación de políticas, la coproducción de bienes y servicios, y el monitoreo, control de gestión y evaluación de los resultados de la gestión estatal” [1].

La innovación pública implica crear soluciones y procesos junto a los ciudadanos con el fin de identificar sus necesidades y también, atender sus opiniones. “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” [2]. Estos espacios-fronteras permiten la comunicación entre distintos actores con diferentes conocimientos e intereses para lograr un entendimiento común y aporte ciudadano.

“La noción de gobierno abierto se muestra como un medio para cambiar la manera en la que las instituciones toman decisiones y se relacionan con esta nueva ciudadanía. La finalidad buscada es reestablecer la confianza pública y lograr una mayor robustez democrática, así como gobiernos más eficientes, eficaces y transparentes en apoyo a los procesos de reforma del Estado y modernización de los servicios públicos” [3]. Este concepto se ve reflejado en un cambio cultural, se trata de modificar la manera de interactuar y de participar de los ciudadanos en las decisiones de gobierno público a través de nuevas tecnologías. Este paso es fundamental para avanzar con políticas concretas que permitan un adecuado proceso de transformación en el tratamiento de transparencia, participación y colaboración en el diseño de la gestión pública.

Es importante el fortalecimiento en la gestión pública para la implementación de estrategias, “como experiencias de implementación del Gobierno Abierto se señalan como estrategias convenientes, comenzar con las innovaciones en pequeñas unidades territoriales, aplicándolas a solucionar problemas concretos, delimitados, que produzcan resultados inmediatos. Esto favorece, por un lado, la micro-segmentación de la participación, por medio de la cual el ciudadano selecciona los temas en los que desea participar y por el otro, posibilita la participación asincrónica” [4].

“La gestión de la información en formatos abiertos, comúnmente llamado “datos abiertos”, el cual consiste en el acceso y uso de la información pública por parte de terceros para entregar nuevos servicios a los ciudadanos. Se trata de poner a disposición de la sociedad, de manera libre, gran cantidad de datos procedentes de diferentes organizaciones del ámbito de la administración pública o de aquellos proyectos que han sido financiados con recursos públicos” [5]. Básicamente, la política de datos abiertos busca promover la innovación y transformar la actividad gubernamental para brindar mejores servicios y, además, devolver a los ciudadanos sus propios datos públicos, generando mayores niveles de transparencia.

Para cada conjunto de datos públicos es necesario determinar su valor, su calidad y quién es el responsable de su mantenimiento/actualización. “Los avances tecnológicos en materia del análisis de grandes bases de datos, datos estadísticos y datos abiertos pasan a tener un rol fundamental, creando posibilidades sin precedentes para informar a la sociedad, transformarla y proteger el medio ambiente. De este modo, en la implementación de la Agenda 2030 se deben

aprovechar las oportunidades que ofrece la revolución de los datos, con el fin de apoyar a los países más necesitados de recursos y dar un impulso sin precedentes hacia un nuevo mundo de datos que propicien el cambio” [6]. La Agenda 2030 [7] para el Desarrollo Sostenible es el documento A/RES/69-315 [8] que se aprobó en la Cumbre Mundial sobre el Desarrollo Sostenible realizada en Nueva York en septiembre de 2015. Los Estados Nacionales miembros de la Organización de Naciones Unidas (ONU) aprobaron en dicho documento 17 Objetivos y 169 Metas que deberán ser cumplidos de aquí al 2030. Los temas implicados en estos objetivos y metas apuntan a erradicar el hambre y lograr la seguridad alimentaria; garantizar una vida sana y una educación de calidad; lograr la igualdad de género; asegurar el acceso al agua y la energía sustentable; promover el crecimiento económico sostenido; adoptar medidas urgentes contra el cambio climático; promover la paz; facilitar el acceso a la justicia y fortalecer una alianza mundial para el desarrollo sostenible. Esta iniciativa instala la vinculación de la sostenibilidad ambiental con la inclusión social y atención de las necesidades de los más vulnerables.

A través de los datos abiertos, se permiten establecer ámbitos de colaboración con los ciudadanos. Por ello, para realizar un proceso de apertura serio y útil, es vital que los datos sean claros, completos, confiables y precisos.

A. Niveles de apertura de los datos

Existen distintos parámetros para definir el nivel de apertura de los datos. Tim Berners-Lee [9], el inventor de la Web e iniciador de los Datos Enlazados (Linked Data), sugirió un esquema de clasificación a partir de 5 estrellas, cada una de las cuales representa un escalón de desarrollo en la apertura de los datos, del más simple al más complejo. Cada categoría está representada con una cantidad de estrellas:

- a) *Una estrella*: implica que el contenido web esté disponible (por ejemplo: archivos .pdf o una imagen), pero no permiten la reutilización;
- b) *Dos estrellas*: se brindan como datos estructurados (por ejemplo: planilla de cálculo en lugar de escanear imágenes de una tabla), aquí se puede trabajar sobre los datos y agregar valor, pero se necesita un software propietario para su lectura;
- c) *Tres estrellas*: están disponibles en un formato abierto no propietario (sin necesidad de pagar por un software) y estructurado (por ejemplo: CSV);
- d) *Cuatro estrellas*: se utilizan los URI para denotar datos. Se mantiene una estructura de datos RDF (Resource Description Framework) cuyo objetivo es describir la semántica de la información de una manera que puede ser interpretada por usuarios o bien algoritmos. Se definen dos niveles diferentes: Uno para el esquema, definiendo la estructura de los datos y otro para los propios datos. Esta combinación de datos, permite formar un esquema global;
- e) *Cinco estrellas*: Este nivel es la excelencia, ya que los datos están vinculados a otros datos para proporcionar un contexto. Se pueden identificar más datos (relacionados) mientras se utilizan éstos.

Por todo lo anteriormente dicho, se “puede hablar de un Gobierno electrónico como un medio para realizar las operaciones de forma más eficiente y eficaz y proveer a la sociedad de servicios públicos de calidad y, por otro lado, de un Gobierno abierto que, además, busca promover el

acercamiento con el ciudadano e impulsar el compromiso mediante la apertura de datos e información” [10].

B. Datos Abiertos versus Datos Públicos

Para obtener información es necesario recurrir a un procedimiento de transformación en los datos. La conversión de éstos genera valor agregado, el cual es un actor fundamental para iniciar una etapa de descubrimiento en la sociedad. “Los datos que se pueden reutilizar y redistribuir sin ninguna restricción se denominan datos abiertos” [11]. Estos datos deben estar en formatos digitales, con un modelo estándar abierto. Por otra parte, “no toda la información pública disponible o publicada en la Web es información abierta válida para su reutilización. No sólo se trata de publicar los datos sino que hay que garantizar el acceso a ellos, razón por la que debe recurrirse a formatos digitales, estandarizados y abiertos, siguiendo una estructura clara que permita su comprensión y reutilización” [12].

El programa de datos abiertos en ciencia y tecnología de la República Argentina, a través de la Resolución 640-E/2017 [13], define que, el dato público es cualquier dato generado en el ámbito gubernamental, o que se encuentre bajo su guarda, y que no tenga su acceso restringido por ninguna legislación específica. Mientras que el dato abierto, es el dato público disponibilizado en un medio digital, bajo una licencia abierta y utilizando para ello un formato estándar abierto. Adicionalmente, y para pertenecer a esta categoría, los datos deberán: estar completos, ser primarios, estar actualizados, ser procesables por máquinas, ser susceptibles de tratamiento, no ser discriminatorios, no ser propietarios, no estar sujetos a derechos de autor (patentes, marcas comerciales o regulación de secretos comerciales).

Finalmente, el Ministerio de Ciencia, Tecnología e Innovación de la República Argentina [14], señala que el formato abierto, es el formato de archivo no propietario, cuya especificación esté documentada públicamente o sea de libre conocimiento e implementación y que esté libre de patentes o de cualquier otra restricción legal o económica en cuanto a su utilización. Para ello, es necesaria una licencia abierta, que es el acuerdo de provisión de datos que conceda amplio acceso para que cualquier persona los utilice, los reutilice y los distribuya, estando sujeto a las siguientes condiciones.

C. Proceso de apertura de datos en el mundo

Actualmente, cada vez son más los países que se unen en este nuevo paradigma de diseñar políticas e iniciativas de datos abiertos. El Open Data Impact Map [15], es un proyecto de la Red de Datos Abiertos para el Desarrollo (OD4D) [16], es una base de datos pública de organizaciones que utilizan datos abiertos del gobierno de todo el mundo, se encuentran disponibles públicamente a los que cualquier persona puede acceder y reutilizar de forma gratuita. En la Figura 1, se muestra un mapa con el impacto de datos abiertos, que tiene una base de datos pública de organizaciones que utilizan datos abiertos de todo el mundo. Participan 90 países, de las cuales hay 1615 organizaciones (empresas, organizaciones sin fines de lucro, instituciones académicas y grupos de desarrolladores) que utilizan datos gubernamentales abiertos, para desarrollar productos y servicios, mejorar operaciones, informar estrategias y realizar investigaciones más completas de la demanda de datos abiertos.

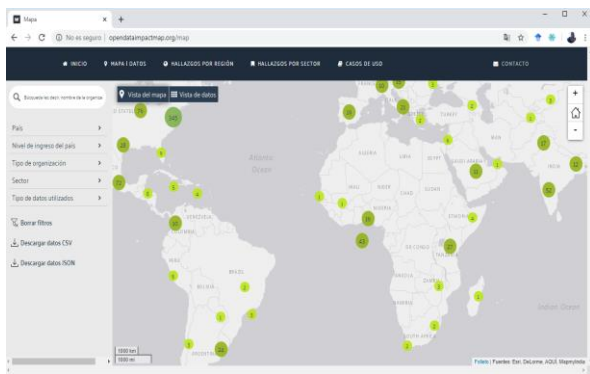


Fig. 1. Mapa de impacto de datos abiertos. Fuente [15].

La fundación Open Knowledge International [17] propone la siguiente metodología de apertura de datos a partir de cuatro pasos: a) *Seleccionar los datos*: analizar los criterios de valor a obtener, los costos para mantenerlos abiertos y riesgos asociados a su publicación; b) *Aplicar una licencia abierta*: los datos deben estar abiertos tanto técnicamente, como legalmente; c) *Abrir los datos*: se realiza a través de los catálogos de datos que están formados por filtros de búsqueda, metadatos y acceso a los conjuntos de datos, básicamente son los portales de datos; d) *Promover el uso*: la comunidad debe conocer su existencia para su reutilización.

Por otra parte, existe un sitio web llamado barómetro de datos abiertos a nivel mundial [18], el cual utiliza una medida global de cómo los gobiernos publican y utilizan datos abiertos para la rendición de cuentas, la innovación y el impacto social. En este sitio se muestra el análisis de 30 gobiernos que han adoptado los principios de datos abiertos en base al Open Data Charter [19]. El barómetro de datos abiertos tiene como objetivo descubrir el verdadero impacto de las iniciativas de datos abiertos en todo el mundo. Analiza las tendencias mundiales y proporciona datos comparativos sobre gobiernos y regiones utilizando una metodología en profundidad que combina datos contextuales, evaluaciones técnicas e indicadores secundarios. Cada uno de estos 30 gobiernos posee una cantidad asignada en base a una clasificación en: a) Disponibilidad para iniciativas de datos abiertos; b) Implementación de programas de datos abiertos; c) Impacto de los datos abiertos en los negocios, la política y la sociedad civil.

En la Figura 2, se muestra el mapa mundial con los 30 gobiernos, discriminados por color en base al puntaje que posee cada país.

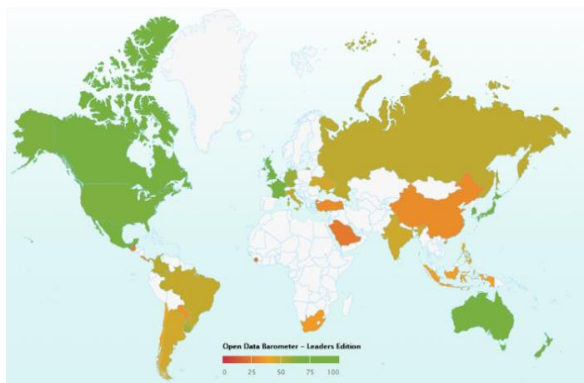


Fig. 2. Calificación de los gobiernos en el sitio del barómetro de datos abiertos. Fuente [18].

Los tonos van desde el rojo (menor puntaje) hasta el color verde (mayor puntaje). Argentina posee 47 puntos (siendo el máximo puntaje a alcanzar 100 puntos), los que mayor puntaje han alcanzado son Canadá y Reino Unido con 76 puntos.

D. Catálogos de datos y conjuntos de datos (datasets)

Los catálogos de datos representan una lista de activos de datos de un organismo del Estado con sus metadatos correspondientes, en éstos se describe que son y cómo pueden ser utilizados. Esto ayuda a encontrar y comprender mejor los datos públicos sobre un tema dado. El concepto de conjunto de datos, más conocido como datasets, se refiere a la parte fundamental en todo el catálogo de datos, es decir, agrupan uno o más recursos de datos. Mantener la calidad en los datasets y sus catálogos, hace que sea posible una correcta accesibilidad para los ciudadanos, por lo que estos conceptos van de la mano.

Esencialmente, la política de apertura de datos de la Administración Pública Nacional se apoya en una red de organismos que publican sus catálogos de datos abiertos siguiendo un estándar internacional de metadatos que facilita su búsqueda y comprensión. Este estándar ayuda y estructura la documentación de activos de datos en datasets.

Existen frameworks que permiten construir los catálogos, por ejemplo: CKAN [20] utilizado por Datos Abiertos Argentina; DKAN [21] utilizado por el Ministerio de Educación de Perú; Socrata [22]: utilizado por la Casa Blanca, Estados Unidos.

II. EXPERIENCIAS DE DATOS ABIERTOS EN ARGENTINA

“Los compromisos institucionales con el Gobierno Abierto por parte de las administraciones públicas y los gobiernos se vertebran a través de diferentes mecanismos políticos y organizativos. Entre dichos mecanismos destacan los denominados planes de acción de Gobierno Abierto, elaborados por aquellas instituciones estatales, regionales o locales que forman parte de la Alianza para el Gobierno Abierto (Open Government Partnership en su denominación en inglés)” [23]. El Open Government Partnership (OGP) [24] es una organización de reformadores dentro y fuera del gobierno, que trabaja para transformar la forma en que el gobierno sirve a sus ciudadanos las cuales generan compromisos a concretar con fecha determinada. En 2011, los líderes gubernamentales y los defensores de la sociedad civil se unieron para crear una asociación única, para promover una gobernabilidad responsable, responsable e inclusiva. Según el sitio web oficial de OGP, existen 68 países y un número creciente de gobiernos locales representando a más de dos mil millones de personas, junto con miles de organizaciones de sociedad civil, los cuales son miembros.

Con respecto a Argentina se incorporó en el año 2012 a la Alianza para el Gobierno Abierto, y actualmente se encuentra en su cuarto Plan de Acción. Para mayor información, se puede consultar la fuente [25], con todo lo que respecta a nuestro País y su compromiso con el Gobierno Abierto. En la República Argentina existe la Ley 104 de acceso a la información pública, en la cual se expone el derecho del ciudadano a obtener información pública [26], esta Ley brinda un reconocimiento al acceso de la información como un derecho humano y, además, formaliza la apertura de datos al ciudadano.

En las siguientes secciones, se analizan los portales de datos abiertos en la República Argentina.

A. Portal de la Ciudad Autónoma de Buenos Aires

En el año 2013 la ciudad autónoma de Buenos Aires publicó el Decreto 478 [27] que estableció como obligatorio que todos los datos producidos, almacenados y/o recolectados en medios digitales sean publicados en formato abierto en la plataforma: data.buenosaires.gob.ar. Existen 354 datasets de 29 organizaciones.

B. Portal Nacional de Argentina

A nivel nacional, la legislación, a través del Decreto 117/2016 [28], estableció la apertura de los datos públicos del Poder Ejecutivo Nacional y su publicación en el portal de datos: datos.gob.ar. En este portal se encuentran todos los datos abiertos de los ministerios, actualmente existen 973 datasets de 33 organizaciones.

C. Portales de Organismos Descentralizados

Se registran los siguientes organismos descentralizados:

- PPN - Procuración Penitenciaria de la Nación: Este portal de la Procuración Penitenciaria de la Nación, pone a disposición información relativa a los derechos de las personas privadas de su libertad según los estándares internacionales de datos abiertos. Existen datasets que tienen por objetivo visibilizar algunas de las principales características estructurales del sistema carcelario argentino que han sido abordadas por el organismo como ejes prioritarios de su trabajo. Los datos expuestos arrojan información acerca de temáticas de especial sensibilidad que permiten comprender las prácticas y dinámicas propias de los espacios de encierro.
- IDERA - Infraestructura de Datos Espaciales: Es un portal con datos abiertos sobre servicios y mapas, orientados a varios temas.
- INDEC - Instituto Nacional de Estadística y Censos: El INDEC ha puesto a disposición de los usuarios bases de datos, abarcando diferentes áreas temáticas para su análisis y procesamiento.
- ARSAT - Empresa Argentina de Soluciones Satelitales: Se muestra información pública de los proyectos y servicios de ARSAT, disponible para que todos los ciudadanos visualicen, descarguen y editen.
- SMN - Servicio Meteorológico Nacional: Antes de visualizar los datos, este portal solicita al usuario, aceptar los términos y condiciones para visualizar los archivos digitales.

En la Tabla 1, se visualiza el análisis de los portales de datos abiertos: a nivel nacional, ciudad autónoma de Buenos Aires y de los organismos descentralizados, realizando una comparativa entre cantidades de datasets ofrecidos, cantidad de organizaciones relacionadas en los datasets, temas y si posee APIs.

En base a la Tabla 1, se identifican varios datasets en los portales gubernamentales en Argentina, la diversidad alarma sobre todo cuando algunos de ellos están clasificados como de 1 estrella, como ser pdf o imágenes. Por supuesto la recomendación sería ofrecer formatos que sean adecuados para la extracción y manipulación de datos, pudiendo llegar al

nivel más alto en estrellas; pero por otra parte no excluir los otros formatos. Es decir, que cada dataset provea distintos formatos incluidos aquellos más avanzados.

TABLA 1. ANÁLISIS DE DATASETS DISPONIBLES EN ARGENTINA. PORTALES NACIONALES Y ORGANISMOS DESCENTRALIZADOS.

Portal	Datasets de la República Argentina				
	Dataset	Temas	Formatos	Org.	API
Argentina unida [29]	973	13	PDF CSV XLSX JSON ZIP SHP GeoJson	33	3
Ciudad Autónoma de Buenos Aires [30]	354	11	CSV PDF XLSX XLS SHP URL JPEG DOCX	29	14
Procuración Penitenciaria de la Nación [31]	5	7	XLS XLSX CSV DOCX	1	0
Infraestructura de Datos Espaciales [32]	14453	25	JSON GML ZIP PNG TITLES KML EXCEL	Varias: No se indican	0
Instituto Nacional de Estadística y Censos [33]	No se indica	8	PDF TXT XLS CSV SAV DBF	Varias: No se indican	0
Empresa Argentina de Soluciones Satelitales [34]	7	3	CSV KML	4	0
Servicio Meteorológico Nacional [35]	9	3	TXT	Datos del clima	0

III. RELEVAMIENTO REALIZADO

Se toma como caso de muestra, Argentina Unida con sus 973 dataset, el motivo de esta elección es que estos datasets están a disposición en distintos formatos abiertos. Se ha analizado los distintos formatos de estos conjuntos de datos, por lo que se pudo observar, que el más utilizado es el tipo CSV con un 61,6%, esto se muestra en la Figura 3, junto con el resto de los porcentajes ofrecidos en esta plataforma para cada uno de sus formatos correspondientes (pdf, csv, xlsx, json, zip, shp, GeoJson). A partir de este análisis, puede deducirse que existe un bajo porcentaje de imágenes y archivos en formato pdf, que son los tipos de archivos que dificultan la manipulación de los datos abiertos.

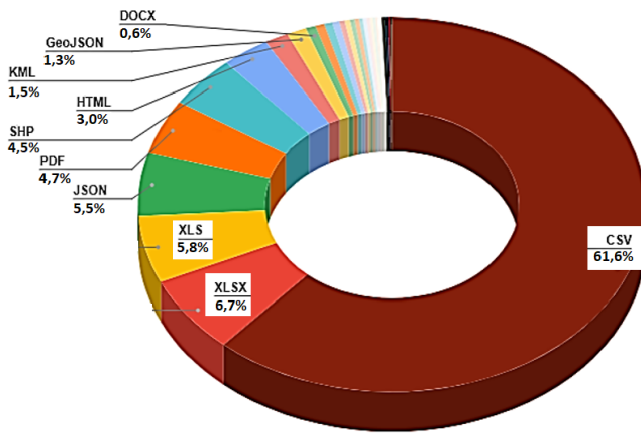


Fig. 3. Formatos de los dataset

A. Portales de Provincias y Municipios

Según el sitio oficial de Portales de datos abiertos de la República Argentina [36], existen varias provincias que tienen portales abiertos: Ciudad de Buenos Aires; Buenos Aires; Catamarca; Córdoba; Entre Ríos; Formosa; Jujuy; Neuquén; Mendoza; Misiones y Tierra del Fuego.

Por otra parte, en el sitio oficial de Portales de datos abiertos de la República Argentina [36], se indican municipios con portales de datos abiertos, los publicados por el Gobierno nacional son: Bahía Blanca; Córdoba; Mar del Plata - Municipalidad de General Pueyrredón; Mercedes; La Plata; Pilar; Rafaela; Tigre; Trenque Lauquen y Villa María.

IV. CATÁLOGOS DE DATOS ABIERTOS

Los organismos de la Administración Pública Nacional deben generar y administrar en forma permanente su propio catálogo incluyendo todos los activos de datos digitales publicados en línea bajo su autoridad, tutela o responsabilidad [37]. El catálogo debe estar publicado en formato JSON o XLSX de estructura compatible. El catálogo debe publicarse en una URL que cumpla alguna de estas 3 condiciones:

- Sea propiedad digital del organismo,
- Esté contenida en el dominio `argentina.gob.ar` o alguno de sus subdominios, bajo el cual el organismo gestione la publicación de sus contenidos y
- Esté contenida en el dominio `datos.gob.ar` o alguno de sus subdominios, bajo el cual el organismo gestione la publicación de datos o metadatos bajo su tutela.

Los organismos de la Administración Pública Nacional deben documentar en su catálogo, todos los activos de datos digitales publicados en línea. Las pautas a cumplir según [37] son:

- Todos los archivos descargables deben estar en formatos: CSV, TXT (tabular), XLS, XLSX, ODS, DTA, SAV, DBF, JSON, XML, GEOJSON, KML, SHP o RDF.
- Todos los archivos de otros formatos no incluidos en los puntos anteriores, siempre deberán sean diseñados para almacenar datos tabulares.
- Todos los archivos de formatos comprimidos ZIP, RAR o cualquier otro que contengan dentro algún archivo de los formatos anteriormente mencionados.

- Todos los archivos PDF, DOC, DOCX, HTML, TXT y otros formatos de documentos que contengan documentación metodológica referida a archivos de los formatos anteriores. Estos se consideran complementarios y deben documentarse siempre en conjunto con los archivos que efectivamente contienen los datos.
- Todas las URLs que contengan documentación de uso de APIs o servicios web de datos total o parcialmente abiertos al público general.

V. PROBLEMAS ENCONTRADOS

Se encontraron algunos problemas en base al análisis realizado en las secciones anteriores, éstos se dividen en a) Portales de Datasets y b) Validaciones de archivos de catálogos de datos abiertos.

A. Portales de Datasets

- No se muestra la cantidad de datasets disponibles en el portal.
- Varios de los temas se encuentran con datasets repetidos, es decir, existe redundancia de contenido. Al dirigirse a un tema, se muestran datasets que se mostraron en otros temas.
- Existen varios formatos que no corresponden a la clasificación de datos abiertos. Por otra parte, varios sitios no poseen una estructura de datos organizada para identificar la información, esto es, el sitio no es amigable para llevar a cabo filtros.
- La mayoría de los sitios no posee APIs para facilitar un mejor nivel de reutilización de datos del portal.
- En varios portales no se registran datasets actualizados.
- En la mayoría de sitios de los datasets, no se pueden previsualizar, es decir, se deben descargar para saber el contenido.
- Algunos de los datasets que se encuentran a descargar, se redireccionan a una URL, como ser: `https://datos.gob.ar/dataset/desarrollo-social-centros-referencia-ministerio-salud-desarrollo-social/archivo/desarrollo-social_124fe88a-9053-4a55-bde8-d3818a3e6abc`, en el que se redirecciona al Ministerio de Desarrollo Social.
- Algunos de los datasets no se pueden descargar, por ejemplo: `http://geoportal.idesa.gob.ar/geoserver/gwc/service/gmaps?layers=geonode:creci_urba2002&zoom={z}&x={x}&y={y}&format=image/png8`; se muestran errores del tipo: 400: For input string: "{z}". Este archivo posee un formato del tipo "titles".
- Varios de los datasets se encuentran en proceso de desarrollo.
- Algunos sitios sólo poseen un único formato de descarga, como ser txt, por ejemplo: sitio del servicio meteorológico.

B. Validaciones de archivos de catálogos de datos abiertos

Para acceder a los datos abiertos es fundamental que estén documentados dentro de un catálogo de datos abiertos. Desde

la Secretaría de Modernización, se incentiva a los distintos organismos a desarrollar el catálogo de datos abiertos. Es por ello que, ofrecen la opción de que los organismos envíen su catálogo, y el equipo de Datos Argentina analiza cada dataset nuevo (antes de ser publicado en el portal del organismo), con el fin de mejorar la calidad de los datos y metadatos. Por otra parte, para los desarrolladores de software, se ofrece pydatajson [38], el cual es un paquete en python con herramientas para generar y validar metadatos de catálogos de datos en formato data.json.

A través del reciente buscador lanzado por Google de datos abiertos, Dataset Search [39], se identifican los catálogos de los portales de la Tabla 1, con el fin de analizarlos, para validar si cumplen con un formato de catálogo de datos, apto para formar parte del sitio gubernamental datos.gob.ar.

En la Tabla 2 se muestra el análisis de la validación de los catálogos de datos de los portales del listado de nodos de la APN según los análisis del perfil nacional de metadatos de datos abiertos en formato .xlsx. También existen en formatos: CSV y JSON. Esta validación se realizó mediante el paquete en Python con la biblioteca necesaria para validar el formato de campos de los catálogos de datos.

TABLA 2. VALIDACIÓN MEDIANTE PYDATAJSON DE LOS CATÁLOGOS DE DATOS ABIERTOS DE LOS PORTALES DEL PERFIL NACIONAL (MINISTERIOS).

Jurisdicción	Catálogos de los portales del perfil nacional https://monitoreo.datos.gob.ar/nodos.xlsx		
	Ente	Link	¿Formato Válido?
Secretaría de Gobierno de Energía	Ente Nacional Regulador del Gas (ENARGAS)	https://transparencia.enargas.gob.ar/data.json	SI
Jefatura de Gabinete de Ministros	Ente Nacional de Comunicaciones (Enacom)	https://datosabiertos.enacom.gob.ar/data.json	NO
Jefatura de Gabinete de Ministros	Jefatura de Gabinete de Ministros	https://monitoreo.datos.gob.ar/catalog/igj/data.json	SI
Jefatura de Gabinete de Ministros	Agencia de Acceso a la Información Pública	https://monitoreo.datos.gob.ar/catalog/aaip/data.json	SI
Secretaría de Gobierno de Modernización	Sin organización asignada	http://monitoreo.datos.gob.ar/catalog/otros/data.json	SI
Secretaría de Gobierno de Modernización	Secretaría de Gobierno de Modernización	https://monitoreo.datos.gob.ar/catalog/modernizacion/data.json	SI
Ministerio de Defensa	Instituto Geográfico Nacional	https://monitoreo.datos.gob.ar/catalog/ign/data.json	SI
Ministerio de Defensa	Servicio Meteorológico Nacional	https://monitoreo.datos.gob.ar/catalog/smn/data.json	SI
Ministerio de Defensa	Ministerio de Defensa	http://datos.mindef.gov.ar/data.json	SI
Ministerio de Justicia y Derechos Humanos	Ministerio de Justicia y Derechos Humanos	https://monitoreo.datos.gob.ar/catalog/justicia/data.json	SI
Ministerio de Seguridad	Ministerio de Seguridad	https://monitoreo.datos.gob.ar/catalog/seguridad/data.json	SI
Ministerio de Transporte	Ministerio de Transporte	https://servicios.transporte.gob.ar/gobierno_abierto/data.json	SI

Jurisdicción	Catálogos de los portales del perfil nacional https://monitoreo.datos.gob.ar/nodos.xlsx		
	Ente	Link	¿Formato Válido?
Ministerio de Transporte	Secretaría de Planificación del Transporte	https://datos.transporte.gob.ar/data.json	SI
Ministerio de Agricultura, Ganadería y Pesca	Ministerio de Agricultura, Ganadería y Pesca	https://datos.agroindustria.gob.ar/data.json	NO
Ministerio de Ambiente y Desarrollo Sostenible	Ministerio de Ambiente y Desarrollo Sostenible	http://datos.ambiente.gob.ar/data.json	SI
Ministerio de Ciencia, Tecnología e Innovación	Ministerio de Ciencia, Tecnología e Innovación	http://datasets.datos.mincyt.gob.ar/data.json	SI
Ministerio de Cultura	Ministerio de Cultura	https://datos.cultura.gob.ar/data.json	SI
Ministerio de Desarrollo Social	Ministerio de Desarrollo Social	https://datosabiertos.desarrollosocial.gob.ar/data.json	SI
Ministerio de Educación	Secretaría de Políticas Universitarias	http://andino.siu.edu.ar/data.json	SI
Ministerio de Educación	Ministerio de Educación	https://monitoreo.datos.gob.ar/catalog/educacion/data.json	SI
Ministerio de Economía	Subsecretaría de Programación Macroeconómica	https://infra.datos.gob.ar/catalog/sspm/data.json	NO
Ministerio de Economía	Subsecretaría de Programación Microeconómica	https://monitoreo.datos.gob.ar/catalog/sspmi/data.json	SI
Ministerio de Economía	Subsecretaría de Presupuesto	https://www.presupuestoabierto.gob.ar/sici/rest-api/catalog/public	NO
Ministerio de Desarrollo Productivo	Secretaría de Transformación Productiva	https://monitoreo.datos.gob.ar/catalog/siep/data.json	SI
Ministerio de Desarrollo Productivo	Secretaría de Energía	http://datos.mine.m.gob.ar/data.json	SI
Ministerio de Desarrollo Productivo	Ministerio de Desarrollo Productivo	https://datos.produccion.gob.ar/data.json	NO
Ministerio de Relaciones Exteriores, Comercio Internacional y Culto	Ministerio de Relaciones Exteriores, Comercio Internacional y Culto	https://monitoreo.datos.gob.ar/catalog/exterior/data.json	SI
Ministerio de Salud	Ministerio de Salud	http://datos.salud.gob.ar/data.json	SI
Ministerio de Salud	Programa de Atención Médica Integral (Pami)	http://datos.pami.org.ar/data.json	NO
Ministerio de Turismo y Deportes	Ministerio de Turismo y Deportes	http://datos.yvera.gob.ar/data.json	SI
Ministerio del Interior	Ministerio del Interior	https://datos.mininterior.gob.ar/data.json	SI
Ministerio de Obras Públicas	Autoridad Cuenca Matanza Riachuelo (Acumar)	http://andino.acumar.gov.ar/data.json	NO

En la Tabla 3, se muestra el análisis de la validación de los catálogos de datos de los portales de la Tabla 1, mediante la biblioteca de Python para validar el formato de campos de los catálogos de datos abiertos. Se puede observar que no todos los sitios ponen a disposición el catálogo de datos abiertos, lo que dificulta una articulación de metodología, al momento de visualizar los datos abiertos. Por otra parte, no todos cumplen con un formato válido.

Como se puede observar en la Tabla 3, de los 7 portales analizados, existen sólo 3 organismos que poseen sus catálogos de datos publicados en formato json y que cumplen con el estándar establecido por el estado nacional argentino.

TABLA 3. VALIDACIÓN MEDIANTE PYDATAJSON DE LOS CATÁLOGOS DE DATOS ABIERTOS DE LOS PORTALES DE LA TABLA 1.

Portal	Catálogos de datos abiertos	
	Link: Catálogo	¿Formato Válido?
Argentina Unida	https://datos.gob.ar/data.json	SI
Ciudad Autónoma de Buenos Aires	https://data.buenosaires.gob.ar/data.json	SI
Procuración Penitenciaria de la Nación	http://datos.ppn.gov.ar/data.json	NO
Infraestructura de Datos Espaciales	No publicado	No publicado
Instituto Nacional de Estadística y Censos	No publicado	No publicado
Empresa Argentina de Soluciones Satelitales	https://datos.arsat.com.ar/data.json	SI
Servicio Meteorológico Nacional	http://datos.gob.ar/superThemeTaxonomy.json	NO

VI. CONCLUSIONES

Los datos abiertos deben estar disponibles para que los ciudadanos puedan acceder a ellos. Es primordial, que estos conjuntos de datos puedan ser identificados con facilidad, en un portal de datos que genere valor a partir de ellos, por lo que, como ya se vio en este trabajo, la importancia del formato brindado y la calidad del dato es fundamental. Los datos necesitan ser abiertos, ya que los ciudadanos deben poder acceder libremente a datos e información del gobierno y compartir esa información con otros. Parte del concepto de transparencia, implica esto, es decir, no sólo de acceder a los datos, sino que además se pueda compartir y reutilizar.

En estos tiempos, los datos son un recurso indispensable para cualquier actividad de gestión pública, es por ellos que, es necesario mantenerlos actualizados, cuidando su calidad. Parte de la participación y compromiso del estado nacional, implica interactuar con los ciudadanos con el fin de conocer sus necesidades.

Producto del relevamiento realizado se pudieron identificar diversos problemas en cuanto a la mejora continua en la calidad de los datasets y catálogos de datos abiertos. Existen varios conjuntos de datos en los que la información no se encuentra actualizada, como así también, pocas alternativas de formatos abiertos para que los datos puedan ser reutilizados. Finalmente, es importante trabajar en la visualización del contenido en un portal amigable, ya que la mayoría de los conjuntos de datos, deben ser descargados para ser visualizados, es decir, sería interesante que todos los

portales cuenten con una previsualización del dataset antes de su descarga.

Este artículo muestra que es un gran campo de acción el de análisis calidad de los datasets, por lo cual continuaremos trabajando en esta línea, con el fin de abordar los parámetros que deben ser considerados al momento de publicar un dataset y desarrollando una automática de validación para el análisis de la calidad de los mismos.

REFERENCIAS

- [1] Oszlak, Oscar, "Políticas públicas y capacidades estatales", Revista Forjando, 2014, vol. 3, no 5.
- [2] Rodríguez, Exequiel, "Laboratorios de innovación pública. Estado Abierto", Revista sobre el Estado, la administración y las políticas públicas, 2018, vol. 3, no 1.
- [3] Acevedo, Sebastián; Dassen, Nicolás, "Innovando para una mejor gestión. La Contribución de los laboratorios de Innovación Pública", Washington: Banco Interamericano de Desarrollo, 2016.
- [4] Gorrochategui, Nora, "Innovación: gobernanza reflexiva y tecnologías de la información y las comunicaciones (TIC) en la gerencia pública Argentina", Trilogía Ciencia Tecnología Sociedad, 2015, vol. 7, no 13, p. 35-46.
- [5] Naser, Alejandra; Ramírez Alujas, Álvaro, "Plan de gobierno abierto: una hoja de ruta para los gobiernos de la región", 2017.
- [6] Montero, Gregorio, "Del gobierno abierto al Estado abierto: la mirada del Centro Latinoamericano de Administración para el Desarrollo. Desde el gobierno abierto al Estado abierto en América Latina y el Caribe", Santiago: CEPAL, 2017. LC/PUB. 2017/9-P. p. 53-81, 2017.
- [7] Agenda 2030 – ODS Argentina. Consejo Nacional de Coordinación de Políticas Sociales. Presidencia de la Nación, "Agenda 2030", Disponible en: <https://www.odsargentina.gob.ar/Agenda2030>, consultado marzo 2020.
- [8] Naciones Unidas. Asamblea General, "Resolución aprobada por la Asamblea General el 1 de septiembre de 2015", Disponible en: <https://undocs.org/es/A/RES/69/315>, consultado marzo 2020.
- [9] 5 Open Data, "5 Open Data", Disponible en: <https://5stardata.info/en/>, consultado marzo 2020.
- [10] Quintanilla, G., & Gil-García, J. R., "Gobierno abierto y datos vinculados: conceptos, experiencias y lecciones con base en el caso mexicano", Revista del clad Reforma y Democracia, 2016, (65), 69-102.
- [11] E. Oviedo, JN Mazón y JJ Zubcoff, "Hacia un modelo de calidad de datos para portales de datos abiertos", 2013 XXXIX Latin American Computing Conference (CLEI), Naiguata, 2013, pp. 1-8.
- [12] Garriga-Portolà, Marc, "¿Datos abiertos? Sí, pero de forma sostenible", El profesional de la información, 2011, vol. 20, no 3.
- [13] Boletín Oficial de la República Argentina. Legislación y Avisos Oficiales, "Ministerio de Ciencia, Tecnología e Innovación Productiva. Resolución 640-E/2017", Disponible en: <https://www.boletinoficial.gob.ar/detalleAviso/primera/167874/20170724>, consultado marzo 2020.
- [14] Argentina.gob.ar, "Ministerio de Ciencia, Tecnología e Innovación de la República Argentina", Disponible en: <https://www.argentina.gob.ar/ciencia>, consultado marzo 2020.
- [15] Open Data Impact Map, "Open Data Impact Map: A public database of organizations that use open data from around the world", Disponible en: <https://www.opendataimpactmap.org/>, consultado marzo 2020.
- [16] ODD, "Open Data for Development", Disponible en: <https://www.od4d.net/about.html>, consultado marzo 2020.
- [17] Open Knowledge International, "How to open data. A quick primer on how data holders can open up their data. How to start small but think big", Disponible en: <https://okfn.org/opendata/how-to-open-data/>, consultado marzo 2020.
- [18] Open Data Barometer – World Wide Web Foundation, "The Open Data Barometer", Disponible en: https://opendatabarometer.org/?_year=2017&indicator=ODB, consultado en marzo 2020.
- [19] The International Open Data Charter, "Open Data Charter", Disponible en: <https://opendatacharter.net/>, consultado en marzo 2020.

- [20] CKAN, “CKAN, the world’s leading Open Source data portal platform”, Disponible en: <https://ckan.org/>, consultado marzo 2020.
- [21] DKAN, “DKAN Open Data Platform”, Disponible en: <https://getdkan.org/>, consultado marzo 2020.
- [22] Socrata, “Building something using data?”, Disponible en: <https://dev.socrata.com/>, consultado marzo 2020.
- [23] Criado, J. I., & Pastor, V., “La Educación en los Planes de Acción de Gobierno Abierto”, RIESED-Revista Internacional de Estudios sobre Sistemas Educativos, 2019, 2(9), 279-300.
- [24] Open Government Partnership, “Sobre Open Government Partnership”, Disponible en: <https://www.opengovpartnership.org/es/about/>, consultado marzo 2020.
- [25] Open Government Partnership, “Argentina: Miembro desde 2012 | Plan de Acción 4”, Disponible en: <https://www.opengovpartnership.org/es/members/argentina/>, consultado marzo 2020.
- [26] La Legislatura de la Ciudad Autónoma de Buenos Aires sanciona con fuerza de Ley, “Ley 104 de acceso a la información pública”, Disponible en: https://www.buenosaires.gob.ar/sites/gcaba/files/ley_5784_-_reforma_ley_104.pdf, consultado marzo 2020.
- [27] Boletín Oficial de la Ciudad de Buenos Aires, “Decreto 478 2013”, Disponible en: <https://boletinoficial.buenosaires.gob.ar/normativaba/norma/234859>, consultado marzo 2020.
- [28] InfoLeg – Ministerio de Economía y Finanzas Públicas, “Decreto 117/2016. Plan de Apertura de Datos”, Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/257755/norma.htm>, consultado marzo 2020.
- [29] Argentina unida, “Datos Argentina”, Disponible en: <https://datos.gob.ar/>, consultado marzo 2020.
- [30] Buenos Aires Ciudad, “Buenos Aires Data”, Disponible en: <https://data.buenosaires.gob.ar/>, consultado en marzo 2020.
- [31] Procuración penitenciaria de la Nación, “Datos PPN”, Disponible en: <http://datos.ppn.gov.ar/>, consultado en marzo 2020.
- [32] IDERA, “Portal IDERA”, Disponible en: <http://catalogo.idera.gob.ar/geonetwork/srv/spa/catalog.search#/home>, consultado en marzo 2020.
- [33] Instituto Nacional de Estadísticas y Censos República Argentina, “Base de Datos”, Disponible en: <https://www.indec.gob.ar/indec/web/Institucional-Indec-BasesDeDatos>, consultado en marzo 2020.
- [34] ARSAT – Presidencia de la Nación, “Portal de datos abiertos de ARSAT”, Disponible en: <https://datos.arsat.com.ar/>, consultado en marzo 2020.
- [35] Servicio Meteorológico Nacional, “Descarga del Catálogo de Datos Abiertos del SMN”, Disponible en: <https://www.smn.gob.ar/descarga-de-datos>, consultado en marzo 2020.
- [36] Portal de datos abiertos- Argentina, “Portales de datos abiertos”, Disponible en: <https://www.argentina.gob.ar/ontti/software-publico/catalogo/datos>, consultado en marzo 2020.
- [37] Perfil de Aplicación Nacional de Metadatos para Datos Abiertos. Secretaría de Modernización. Presidencia de la Nación, “Perfil de Aplicación Nacional de Metadatos para Datos Abiertos”, Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/perfil-metadatos/#condiciones-de-cumplimiento-del-perfil>, consultado marzo 2020.
- [38] Manual de uso –Documentación de pydatajson, “Validar los metadatos de un catálogo y corregir errores.”, Disponible en: <https://pydatajson.readthedocs.io/es/stable/MANUAL.html#validacion>, consultado marzo 2020.
- [39] Dataset Search, “Dataset Search”, Disponible en: <https://datasetsearch.research.google.com/>, consultado marzo 2020.