



Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales

Tesista: Mg. Ing. Roxana Martínez

Directora: Dra. Rocío Rodríguez

Codirectora: Dra. Claudia Pons

Asesor Científico: Dr. Pablo Vera

*Tesis presentada para obtener el título de
Doctor en Ciencias Informáticas*



Septiembre 2021



Resumen

En estos tiempos, los datos son un recurso indispensable para cualquier actividad de gestión pública, por lo que es necesario mantenerlos actualizados, claros y completos. Si bien se puede determinar que cada vez son más los gobiernos que se embarcan en este nuevo paradigma de concepto de datos abiertos, todavía falta un largo camino por recorrer. Varios trabajos relacionados en esta temática se están enfocando en la calidad de los datos abiertos disponibilizados en los portales de entidades gubernamentales, este especial foco en la calidad del dato, permite que se estudien técnicas para que sean mejor utilizados, manipulados y distribuidos a los ciudadanos para varios fines, permitiendo a su vez, una mayor participación ciudadana y transparencia en organismos estatales. Este trabajo se enfoca en el contexto de Gobierno Abierto para colaborar con el cumplimiento, divulgación y mejora de los distintos criterios y principios en aspectos del tratamiento del dato público y abierto que proponen entidades mundialmente reconocidas

Con el fin de identificar mejoras en los aspectos de calidad de los datasets abiertos, esta tesis plantea métricas críticas y no críticas para su análisis y validación de contenido, por lo que, como parte de la propuesta, se presenta un prototipo de desarrollo propio, llamado HEVDA (*HErramienta de Validación de Datos Abiertos*). A modo de caso de estudio, se extrae una muestra de datasets públicos estatales que son validados con HEVDA, para obtener un análisis sobre las mediciones utilizadas y realizar así, un estudio cuantitativo sobre los resultados arrojados. Esta herramienta de validación permite detectar en forma más sencilla y rápida, las falencias y errores en las fuentes de datos abiertas que podrían complicar la interoperabilidad para su utilización en diversos orígenes de bases de datos y softwares externos de otros organismos. Para evaluar la calidad de datos es necesario tener en cuenta determinadas características en el conjunto de datos analizados, por lo que se realiza un relevamiento detallado de los aspectos más relevantes en cuestiones de calidad de datos sobre criterios estándares de norma ISO/IEC 25012 [ISO08], estándares universales de calidad de datos [CA15], dimensiones de la calidad de los datos [DIM18], trabajos relevados y estudios realizados [MAR20] en esta temática.

La utilización de métricas de calidad favorece al encuadre de indicadores que permitan obtener un dato más limpio para facilitar el análisis final. Con los resultados obtenidos en el estudio realizado de la presente tesis, se puede concluir que sí 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 y sencilla.

Palabras claves:

Datos Abiertos, Datos Públicos, Gobierno Abierto, Métricas de Calidad de Datos.





Dedicatoria

*Dedicado a mi querido hijo, Sebastián, por su gran apoyo incondicional,
cariño y fuerza en todo momento;
Sos mi orgullo.*

*A mi madre María Catena y mi padre José Manuel, quienes me brindaron
las herramientas de vida y me demostraron que, a pesar de las
contingencias, siempre hay que seguir adelante;
Abrazo al cielo para los dos.*





Agradecimientos

- A la Universidad Abierta Interamericana (UAI), especialmente al Dr. Marcelo De Vincenzi, Dr. Carlos Neil y Dra. Claudia Pons por el apoyo, seguimiento y confianza en mí para llevar a cabo este nuevo desafío, el doctorado.
- A mi directora de tesis Dra. Rocío Rodríguez, a mi Codirectora Dra. Claudia Pons y a mi asesor técnico de tesis Dr. Pablo Vera, por volver a confiar en mí en este nuevo posgrado, y quiénes me ofrecieron su experiencia y contención para culminar mi tesis.
- Al personal de la Universidad Nacional de La Plata, autoridades, personal administrativo, docentes, etc. por brindarme el espacio para realizar este posgrado, y por su trato tan cordial cada vez que tuve una duda o consulta.
- Al personal de la Subsecretaría Gobierno Abierto de la Ciudad Autónoma de Buenos Aires, por las respuestas brindadas a las consultas efectuadas y cursos sugeridos a lo largo de esta tesis.
- Al personal del Instituto Nacional de la Administración Pública (INAP) por las respuestas ofrecidas en consultas y apoyo en cursos brindados por la Administración Pública nacional (APN).
- Al personal de la Oficina Nacional de Tecnologías de Información (ONTI) encargada de dirigir la formulación de políticas e implementación del proceso de desarrollo e innovación tecnológica para la transformación e innovación del Estado Nacional, por la información otorgada frente a dudas y consultas. Especialmente a los asesores técnicos Hernán Mavrommatis y Pablo Ferrante.
- A Carolina Soleil, Susana Soto y Mariano Grimoldi, personal del Centro de Altos Estudios en Tecnología Informática (CAETI) y Biblioteca, por el apoyo brindado y colaboración en tareas administrativas de investigación y presentación en Congresos.
- A mis colegas de trabajo de la oficina de la Dirección de Seguridad y Sistemas de la Información (DSSI) de la Unidad de Información Financiera (UIF) de la Nación.
- A mis amigos, colegas y estudiantes, quienes me apoyaron en este largo camino y aprendo cada día.
- Un especial agradecimiento a queridos colegas y amigos del entorno académico, que siempre recuerdo con cariño: Santiago Sábato, Jorge Scali, Fernando Asteasuain, Gabriela Arévalo, Javier Bazzocco, Alejandro Fernández y Diego Torres. Un agradecimiento especial a Luciano Straccia con quién compartí tantos años de trabajo en diferentes universidades.
- A mis colegas de varios cursos de este posgrado en Ciencias Informáticas en la UNLP, tanto nacionales como internacionales, de quienes compartimos varios trabajos y ricas experiencias: Glenda, Jorge, Edith, Gladys, Héctor, Eduardo, Ronny, Jonathan y Franklin.





Índice General

Capítulo 1 - Introducción.....	- 17 -
1.1. Contexto	- 17 -
1.2. Definición del Problema	- 21 -
1.3. Objetivo General	- 24 -
1.4. Objetivos Específicos	- 24 -
1.5. Hipótesis	- 25 -
1.6. Alcance.....	- 25 -
1.7. Contribución Científica	- 25 -
1.8. Enfoque Metodológico.....	- 26 -
1.9. Organización de la Tesis	- 28 -
Capítulo 2 - Marco Teórico	- 30 -
2.1. Innovación Pública	- 30 -
2.2. Gobierno Abierto.....	- 31 -
2.3. Datos Públicos	- 32 -
2.3.1. Acceso a la Información Pública	- 33 -
2.3.1.1. Transparencia Activa	- 34 -
2.3.1.2. Transparencia Pasiva.....	- 34 -
2.4. Datos Abiertos	- 34 -
2.4.1. Niveles de Apertura de Datos	- 35 -
2.4.2. Apertura de datos en el mundo.....	- 36 -
2.4.3. Apertura de datos en América Latina y el Caribe.....	- 39 -
2.4.4. Apertura de datos abiertos en Argentina.....	- 40 -
2.4.4.1. Relevamiento de Portales abiertos gubernamentales provinciales argentinos .	- 42 -
2.5. Catálogos de datos y conjuntos de datos (datasets).....	- 46 -
2.6. Principios internacionales de los datos abiertos.....	- 48 -
2.7. Aspectos de Calidad: Norma y Estándares	- 48 -
Capítulo 3 – Trabajos Relacionados	- 52 -
3.1. Construcción de Bancos de datos abiertos.....	- 53 -
3.2. Evaluación de la reutilización de datos abiertos	- 57 -
3.3. Crowdsourcing y datos abiertos	- 58 -



3.4. Calidad en los portales de datos abiertos	- 61 -
3.5. Medición de calidad	- 63 -
3.6. Herramientas de validación de datos abiertos	- 64 -
3.7. Resumen de Trabajos relevados	- 69 -
Capítulo 4 – Definición de Métricas para el análisis de Datasets Gubernamentales	- 71 -
4.1. Métricas Críticas	- 72 -
4.1.1. Números Decimales.....	- 72 -
4.1.2. Registros Duplicados	- 74 -
4.1.3. Datos Faltantes y/o Completos.....	- 76 -
4.1.4. Caracteres Inválidos	- 80 -
4.2. Métricas No Críticas	- 81 -
4.2.1. Redundancia para el dominio de una columna.....	- 81 -
4.2.2. Redundancia entre campos de una misma fila	- 83 -
4.2.3. Detección de valores ID.....	- 85 -
4.2.4. Campos Triviales.....	- 87 -
Capítulo 5 – Implementación	- 89 -
5.1. Herramienta Desarrollada - HEVDA.....	- 89 -
5.1.1. Aspectos Técnicos.....	- 89 -
5.1.2. Aspectos Funcionales	- 90 -
5.2.1. Métricas	- 93 -
5.2.1.1. Tratamiento de Números Decimales.....	- 93 -
5.2.1.2. Registros Duplicados	- 94 -
5.2.1.3. Datos Faltantes y/o Completos	- 95 -
5.2.1.4. Caracteres Inválidos	- 97 -
5.2.1.5. Redundancia para el dominio de una columna.....	- 98 -
5.2.1.6. Redundancia entre campos de una misma fila	- 99 -
5.2.1.7. Detección de valores ID.....	- 101 -
5.2.1.8. Campos Triviales.....	- 102 -
5.3. Recolección de datos.....	- 102 -
5.4. Análisis de resultados.....	- 104 -
5.4.1. Análisis de las métricas bloqueantes.....	- 104 -
5.4.2. Resultados generales de las métricas.....	- 105 -
Capítulo 6 – Validación	- 111 -
6.1. Introducción.....	- 111 -



6.2. Aspectos de Validación	- 112 -
6.2.1. Precisión	- 114 -
6.2.2. Exactitud	- 115 -
6.2.3. Completitud	- 116 -
6.2.4. Integridad	- 117 -
6.2.5. Consistencia	- 117 -
6.2.6. Relación entre valores de campos	- 118 -
6.2.7. Estructurales/Representación	- 119 -
6.2.8. Redundancia	- 120 -
6.3. Análisis de los criterios aplicados a las métricas sugeridas	- 121 -
Capítulo 7 – Conclusiones y Trabajo Futuro	- 123 -
7.1. Reflexiones Finales	- 123 -
7.2. Ejecución de los Objetivos.....	- 125 -
7.3. Líneas Futuras de Investigación	- 127 -
7.4. Publicaciones en Revistas y Congresos	- 128 -
Anexos	- 135 -
Anexo A: ISO/IEC 25012	- 135 -
Anexo B: Principios Internacionales de los datos abiertos	- 137 -
Anexo C: Funciones secundarias de la Herramienta HEVDA.....	- 139 -
Anexo D: Portales de datos abiertos gubernamentales.....	- 141 -
Anexo E: Detalle de los resultados arrojados con HEVDA	- 144 -
Anexo F: Detalle de los criterios de Validación	- 181 -
Acrónimos.....	- 185 -
Referencias Bibliográficas	- 189 -





Índice de Figuras

<i>Figura 1.1 – Ecosistema de Gobierno Abierto.</i>	<i>- 18 -</i>
<i>Figura 1.2 – 5 Ejes – Estrategia de Modernización del Estado.</i>	<i>- 21 -</i>
<i>Figura 1.3 – Enfoque Metodológico de la Tesis Doctoral.</i>	<i>- 27 -</i>
<i>Figura 2.1 –Esquema de evolución por niveles de los formatos abiertos para las 5 estrellas.</i>	<i>- 36 -</i>
<i>Figura 2.2 – Mapa de impacto de datos abiertos.</i>	<i>- 38 -</i>
<i>Figura 2.3 – Calificación de los gobiernos en el sitio del barómetro de datos abiertos.</i>	<i>- 39 -</i>
<i>Figura 2.4 – Modelo de Calidad definido por el estándar ISO/IEC 25012. Fuente [ISO12].</i>	<i>- 50 -</i>
<i>Figura 3.1 – Prototipo QILOGD. Pantalla de vista de conjunto de datos.</i>	<i>- 64 -</i>
<i>Figura 3.2 – Prototipo. Pantalla de ingreso de estructura de datos.</i>	<i>- 65 -</i>
<i>Figura 3.3 – Prototipo. Pantalla de validación de calidad de datos abiertos.</i>	<i>- 65 -</i>
<i>Figura 3.4 – Pantalla de la herramienta OHDSI ATLAS.</i>	<i>- 66 -</i>
<i>Figura 3.5 – Pantalla del Prototipo para medir la calidad de datos con métricas.</i>	<i>- 66 -</i>
<i>Figura 3.6 – Pantalla de la herramienta de validación de metadatos INSPIRE.</i>	<i>- 67 -</i>
<i>Figura 3.7 – Una vez completado el proceso de validación, se muestran los errores encontrados.</i>	<i>- 67 -</i>
<i>Figura 3.8 – Validación de calidad de datos abiertos de Argentina, con nivel de apertura.</i>	<i>- 68 -</i>
<i>Figura 3.9 – Herramienta de revisión de datos beta OCDS.</i>	<i>- 69 -</i>
<i>Figura 4.1 – Conjunto de datos para el análisis de Números Decimales.</i>	<i>- 73 -</i>
<i>Figura 4.2 – Dataset con Registros Duplicados (Registros 1 al 17).</i>	<i>- 75 -</i>
<i>Figura 4.3 – Dataset con Registros Duplicados (Registros 1101 al 1117).</i>	<i>- 76 -</i>
<i>Figura 4.4 – Recomendaciones sobre agrupación de conceptos.</i>	<i>- 78 -</i>
<i>Figura 4.5 – Datasets para la métrica de datos faltantes y/o completos.</i>	<i>- 79 -</i>
<i>Figura 4.6 – Dataset para la Métrica de caracteres especiales.</i>	<i>- 80 -</i>
<i>Figura 4.7 – Buenas prácticas de las descripciones de datos abiertos.</i>	<i>- 82 -</i>
<i>Figura 4.8 – Pantalla de Medición de Redundancia en el dominio de valores de una columna.</i>	<i>- 83 -</i>
<i>Figura 4.9 – Dataset para Redundancia entre campos de una misma fila (registros 1289 al 1301).</i>	<i>- 84 -</i>
<i>Figura 4.10 – Dataset para Redundancia entre campos de una misma fila (registros 1346 al 1358).</i>	<i>- 85 -</i>
<i>Figura 4.12 – Dataset para la detección de casos del tipo ID.</i>	<i>- 87 -</i>
<i>Figura 4.13 – Dataset para el análisis de campos triviales.</i>	<i>- 88 -</i>
<i>Figura 5.1 – Pantalla inicial de la herramienta HEVDA.</i>	<i>- 92 -</i>
<i>Figura 5.2 – Métricas Críticas: Pantalla de Verificación de Números Decimales.</i>	<i>- 93 -</i>
<i>Figura 5.3 – Detalle de los casos detectados para la verificación de números decimales.</i>	<i>- 94 -</i>
<i>Figura 5.4 – Métricas Críticas: Pantalla de Verificación de Registros Duplicados.</i>	<i>- 94 -</i>
<i>Figura 5.5 – Detalle de los casos detectados para la verificación de registros duplicados.</i>	<i>- 95 -</i>
<i>Figura 5.6 – Pantalla para la detección de datos faltantes y/o completos.</i>	<i>- 96 -</i>
<i>Figura 5.7 – Pantalla de detalle de casos para detección de datos faltantes.</i>	<i>- 97 -</i>
<i>Figura 5.8 – Pantalla de verificación de caracteres inválidos y la cantidad de casos por columnas.</i>	<i>- 98 -</i>
<i>Figura 5.9 – Detalle de los registros encontrados para los caracteres inválidos.</i>	<i>- 98 -</i>
<i>Figura 5.10– Pantalla de Medición de Redundancia en el dominio de valores de una columna.</i>	<i>- 99 -</i>
<i>Figura 5.11– Detalle - Medición de la redundancia en el dominio de valores de una columna.</i>	<i>- 99 -</i>
<i>Figura 5.12– Pantalla de Medición de Redundancia entre campos de una misma fila.</i>	<i>- 100 -</i>
<i>Figura 5.13– Detalle para la medición de la redundancia entre campos de una misma fila.</i>	<i>- 101 -</i>
<i>Figura 5.14– Métrica de análisis de ID.</i>	<i>- 101 -</i>
<i>Figura 5.15– Métrica de campos triviales.</i>	<i>- 102 -</i>
<i>Figura 5.16 – Porcentaje de Tipos de bloqueantes y no bloqueantes detectados.</i>	<i>- 104 -</i>
<i>Figura 5.17 – Verificación de métricas de calidad de datos abiertos.</i>	<i>- 107 -</i>
<i>Figura 5.18 – Porcentaje de casos encontrados con incumplimiento de métricas.</i>	<i>- 108 -</i>
<i>Figura 5.19 – Porcentaje de casos con incumplimiento de métricas críticas.</i>	<i>- 109 -</i>
<i>Figura 5.20 – Porcentaje de casos con incumplimiento de métricas no críticas.</i>	<i>- 110 -</i>
<i>Figura 6.1 – Criterios seleccionados del Estándar universal de calidad de datos.</i>	<i>- 113 -</i>
<i>Figura 6.2 – Criterios seleccionados de la Norma ISO/IEC 25012.</i>	<i>- 113 -</i>
<i>Figura 6.3 – Criterios seleccionados de las dimensiones de calidad de datos.</i>	<i>- 114 -</i>
<i>Figura 6.4 – Resumen de criterios de validación.</i>	<i>- 122 -</i>

<i>Figura Anexo C.1 – Pantalla de resumen de la Herramienta HEVDA.</i>	- 139 -
<i>Figura Anexo C.2 – Pantalla de resumen de métrica críticas en HEVDA.</i>	- 140 -
<i>Figura Anexo D.3 – Portal del Ministerio de Salud de Argentina [SAL21].</i>	- 141 -
<i>Figura Anexo D.4 – Portal de Diputados de la Argentina [DIP21].</i>	- 142 -
<i>Figura Anexo D.5 – Portal del Ministerio de Justicia y Derechos Humanos de Argentina [MIN21].</i> ...	- 142 -
<i>Figura Anexo D.6 – Portal de la Ciudad de Buenos Aires [BAC21].</i>	- 143 -
<i>Figura Anexo D.7 – Portal Nacional de Argentina [AUD21].</i>	- 143 -



Índice de Tablas

<i>Tabla 1.1 – Cuadro con los distintos componentes del Ecosistema de Gobierno Abierto.</i>	<i>- 19 -</i>
<i>Tabla 5.1 – Cantidad de datasets para cada uno de los portales gubernamentales.</i>	<i>- 103 -</i>
<i>Tabla 5.2 – Cantidad de datasets totales con bloqueantes.</i>	<i>- 104 -</i>
<i>Tabla 5.3 – Cantidad de datasets por tipo de bloqueantes.</i>	<i>- 105 -</i>
<i>Tabla 5.4 – Cantidad de datasets que cumplen y no cumplen por métrica.</i>	<i>- 106 -</i>
<i>Tabla 6.1 – Criterio de Precisión.</i>	<i>- 115 -</i>
<i>Tabla 6.2 – Criterio de Exactitud.</i>	<i>- 115 -</i>
<i>Tabla 6.3 – Criterio de Completitud.</i>	<i>- 116 -</i>
<i>Tabla 6.4 – Criterio de Integridad.</i>	<i>- 117 -</i>
<i>Tabla 6.5 – Criterio de Consistencia.</i>	<i>- 118 -</i>
<i>Tabla 6.6 – Criterio de relación de valores.</i>	<i>- 118 -</i>
<i>Tabla 6.7 – Criterio de estructura/Representación.</i>	<i>- 119 -</i>
<i>Tabla 6.8 – Criterio de redundancia.</i>	<i>- 120 -</i>
<i>Tabla 6.9 – Criterios de Validación.</i>	<i>- 121 -</i>
<i>Tabla Anexo E.1 – Detalle de datasets validados del Ministerio de Salud.</i>	<i>- 144 -</i>
<i>Tabla Anexo E.2 – Detalle de datasets validados de Diputados de la Nación Argentina.</i>	<i>- 145 -</i>
<i>Tabla Anexo E.3 – Detalle de datasets validados de Justicia Argentina.</i>	<i>- 146 -</i>
<i>Tabla Anexo E.4 – Detalle de datasets validados de la Ciudad de Buenos Aires, Argentina.</i>	<i>- 147 -</i>
<i>Tabla Anexo E.5 – Detalle de datasets validados de la República Argentina.</i>	<i>- 156 -</i>







Capítulo 1 - Introducción

En este capítulo se presenta el contexto introductorio sobre el que se desarrolla este trabajo. Se explica de manera breve los conceptos teóricos de Gobierno Abierto, datos abiertos y datos públicos. Luego, se define el problema, el objetivo principal y los objetivos específicos, como así también la hipótesis, el alcance, la contribución científica y el enfoque metodológico utilizado. Finalmente, se muestra la organización de la estructura de la tesis doctoral.

1.1. Contexto

Actualmente se utilizan las TIC (Tecnologías de la Información y la Comunicación) para la colaboración abierta en cuestiones asociadas a la resolución de diversos problemas o temas ciudadanos, que son llevados a cabo a través de sitios web gubernamentales, el enfoque principal se centra en la colaboración y participación de los ciudadanos y organismos públicos. “Esa composición la transforma en un instrumento único para impulsar la cooperación horizontal, el apoyo a la elaboración de políticas de gobierno digital centradas en el ciudadano, la formación de los funcionarios públicos, fomentar el conocimiento de aspectos claves de la construcción de una transformación digital, y el intercambio de soluciones y expertos entre los países” [BID21].

El Estado Nacional Argentino se encuentra promoviendo la participación ciudadana digital a través de herramientas de software en forma on-line, esta idea parte del concepto de Gobierno Abierto en la que la calidad de la información compartida por las entidades gubernamentales permite “promover y facilitar su uso siendo un instrumento que apoya el cumplimiento de los tres pilares del gobierno abierto: Transparencia, Participación y Colaboración” [OLA18].

El concepto de Gobierno Abierto propone que se trabaje de manera colaborativa sobre algunos puntos FODA (Fortalezas, Oportunidades, Debilidades y Amenazas) con respeto a varias situaciones a nivel Estado Nacional, con el fin que tanto las entidades gubernamentales como los ciudadanos, puedan brindar soluciones específicas a determinados problemas trabajando en conjunto. Si bien, el objetivo principal es la colaboración, existe otro punto a destacar que es el tratamiento de los datos colaborados, los cuales implican que, al ser compartidos, el gobierno es abierto y “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” [W3C09].



Mediante este cambio de paradigma, basado en los principios de transparencia, participación y colaboración, se espera una relación más cercana y afianzada entre los ciudadanos de un país y el propio Estado Nacional.

En la Figura 1.1 (extraída de [ALU12]) puede observarse una propuesta sobre un ecosistema de desarrollo de gobierno abierto que brinda los tres principios y elementos que se relacionan a éstos, como ser:

- Visualización de datos e Información;
- Portales de datos abiertos;
- Gestionar la transversalidad e interoperabilidad;
- Nuevos modelos de gestión de servicios públicos;
- Organizar y promover asociaciones (Estado, sociedad civil y empresas);
- Cocreación de valor público;
- Codiseño de políticas públicas;
- Consultas ciudadanas;
- Organizaciones de debates públicos;



Figura 1.1 – Ecosistema de Gobierno Abierto.

Se proyecta que “la implantación eficaz de un Gobierno abierto requiere de un marco legal, de los instrumentos de política pública, de las instituciones y herramientas apropiadas, conseguirlo depende también de la creación y consolidación de una cultura a favor de la apertura en todo el sector público” [ALU11].



De la mano de Gobierno Abierto, existe la gestión colaborativa que involucra el concepto de crowdsourcing, en el que, mediante un llamado abierto a la resolución de un problema o participación de un tema, los ciudadanos pueden involucrarse brindando posibles alternativas a una situación específica. La gestión colaborativa permite la administración de los datos públicos compartidos, lo que hace necesario que se logre una correcta gestión de las plataformas informáticas implicadas con dichos datos. Esta apertura de datos e información a la comunidad fortalece a la eficiencia en la gestión y optimización de las solicitudes de los ciudadanos.

A modo de ejemplo, se presenta en la Ciudad de Buenos Aires, una iniciativa de transparencia activa, innovación, participación ciudadana, colaboración y rendición de cuentas estatales, permitiendo un enfoque especial en 15 puntos fundamentales que conforman el ecosistema propuesto [GBA21]. En la Tabla 1.1 se muestran los distintos componentes del Ecosistema de Gobierno Abierto.

Tabla 1.1 – Cuadro con los distintos componentes del Ecosistema de Gobierno Abierto.

Ícono	Descripción del componente del Ecosistema de Gobierno Abierto
	Plano Abierto Buenos Aires: presenta información abierta de manera sencilla y actualizada sobre la planificación urbana.
	Presupuesto abierto: El presupuesto refleja lo que hace el Estado con los recursos que obtiene de todos los ciudadanos. Es importante comunicar cómo se utilizan esos fondos para estimular la participación, optimizar los controles y mejorar el nivel de la discusión pública.
	BA Obras: Se muestran las obras públicas realizadas en la ciudad y además, se brinda información para cada una de ésta.
	Participación Ciudadana: Se muestran las distintas actividades entre ciudadanos y el Gobierno, con el fin de proponer mejoras para la ciudad.
	Gestión Colaborativa de Solicitudes: Se pueden realizar solicitudes y consultas sobre distintas categorías (Calles y Veredas, Alumbrado, Pluviales, etc.).
	BA en OGP: Información referida a la implementación de las iniciativas de la Ciudad de Buenos Aires en los Planes de Acción ante la Alianza para el Gobierno Abierto.
	Legado Olímpico: Se muestran los datos abiertos sobre los Juegos Olímpicos Buenos Aires 2018 de la Juventud.
	BA Elige: Se muestran propuestas de ideas para mejorar la calidad de vida en la ciudad.
	Plan Estratégico Participativo BA 2035: Se muestran periódicamente planes estratégicos consensuados que ofrecen fundamentos para la elaboración de políticas de Estado en la Ciudad de Buenos Aires.
	Dialogando BA: Es un programa que a través de mesas temáticas de diálogo aporta a la construcción de instituciones democráticas y políticas sustentables.
	BA Ciudad Colaborativa: Este componente promueve información para la construcción colaborativa entre empresas y el gobierno.
	Observatorio Turístico: El Observatorio genera inteligencia de mercado para orientar la toma de decisiones en el sector público, empresas y profesionales, potenciando el turismo como una actividad estratégica para el desarrollo económico y social de la Ciudad de Buenos Aires.



Ícono	Descripción del componente del Ecosistema de Gobierno Abierto
	Mapa de Oportunidades Comerciales: Se muestran indicadores comerciales con gráficos y probabilidades según indicadores por zonas en la ciudad de Buenos Aires.
	BA Data: Se muestran datasets con datos que produce la Ciudad para desarrollar aplicaciones, contar historias y poner en valor los datos de Buenos Aires. Se discriminan por categorías de temas.
	Compromisos de Gobierno: Se muestran las noticias de los objetivos precisos para construir una mejor ciudad.

Para este tipo de iniciativas, es importante mantener y administrar una plataforma en la que la colaboración por parte de las personas sea eficiente, rápida y segura.

Un ejemplo interesante sobre las tecnologías orientadas a la participación ciudadana es “IdeaScale” [IDE21], esto es “una plataforma web de gestión de ideas que fue utilizada por distintas instituciones públicas, políticos y autoridades en sus intentos de generar espacios alternativos de intercambio de opinión, ideación y deliberación entre ciudadanos. Utilizada por la Casa Blanca para justamente recolectar ideas sobre medidas, estrategias y políticas que permitan al gobierno, llevar una administración más transparente, participativa e inclusiva (OpenGov)” [SAL15]. Mediante una plataforma colaborativa las ideas pueden ser valorizadas y pueden cobrar un cierto poder para alcanzar un determinado nivel de importancia y así ser tratadas con prioridad en una entidad gubernamental.

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 RISP (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.

El paradigma de datos abiertos nacional surge de una política impulsada desde el Gobierno de la Nación Argentina, a través del Decreto N° 117/2016 del 12 de enero de 2016, en el cual se hace hincapié al gran aporte de los datos abiertos en los sitios web gubernamentales, de allí el concepto de conjunto de datos abiertos (datasets). Por otra parte, el Ministerio de Gobierno, impulsa una serie de buenas prácticas, guías, seminarios y recomendaciones varias para las instituciones gubernamentales, con el objetivo de convertir y disponibilizar más datos públicos.

Se considera que “resulta necesario aumentar la calidad de los servicios provistos por el Estado incorporando TIC (Tecnologías de la Información y de las Comunicaciones), simplificando procedimientos, propiciando reingenierías de procesos y ofreciendo al ciudadano la posibilidad de mejorar el acceso por medios electrónicos a información personalizada, coherente e integral” [MOD18].

En la Figura 1.2 se presentan los 5 ejes propuestos en el Decreto 434/2016, los cuales muestran la estrategia de modernización del Estado Argentino [INA17].



Figura 1.2 – 5 Ejes – Estrategia de Modernización del Estado.

Según el Decreto 434/2016, la estrategia Nacional se estructura en cinco ejes: a) Plan de Tecnología y Gobierno Digital; b) Gestión Integral de los Recursos Humanos; c) Gestión por los Resultados y Compromisos Públicos; d) Gobierno Abierto e Innovación Pública; y e) Estrategia País Digital. De estos cinco ejes mencionados, existen dos que priorizan el uso de las TIC, como ser:

1. *Plan de Tecnología y Gobierno Digital*: los cuales pretenden incorporar infraestructura tecnológica y redes que son la condición necesaria para agregar nuevos modelos de gestión basados en la tecnología de Big Data;
2. *Gobierno Abierto e Innovación pública*: el cual no sólo se complementa con el anterior eje fundamental, sino que, además, permite la evaluación y control de las políticas públicas del Estado Nacional. Es decir que, la administración de grandes volúmenes de información es un recurso valioso y necesario.

Por lo que, partiendo de esta premisa, hay una gran necesidad de gestionar esta estructura mediante la utilización e incorporación de nuevas herramientas de análisis de datos.

1.2. Definición del Problema

Varios Gobiernos implementaron diversas tecnologías de gestión para el contexto anteriormente mencionado, sin embargo, la aplicación de éstas y la utilización de la digitalización en diversos procesos que existen hoy, no son una solución completa. El marco teórico-conceptual del que surgen los argumentos por los cuales las administraciones públicas se están reordenando y



rediseñando para asumir ciertos rasgos de la empresa privada moderna, consisten en que se pueden formar equipos de trabajo del conocimiento competitivos que fomenten las TICs como vínculos entre la sociedad y sus administraciones públicas, por lo que, las metodologías para evaluar la relación y el impacto de las TIC en las administraciones públicas actuales aún son incipientes y continúan desarrollándose [AVI14].

La gestión pública actual, requiere de un cambio interno en las pautas culturales, organizacionales y normativas y en su relación con el ciudadano. La agilidad, la innovación y la capacidad de adaptación a cambios constantes en el entorno se convirtieron en nuevos valores públicos. La innovación podría mejorar las actividades del gobierno con las necesidades de los ciudadanos actuales, y ayudar a resolver algunos inconvenientes que se presentan en el diseño e implementación de las políticas públicas de hoy. Uno de los enfoques que proponen los Gobiernos con respecto a estos aspectos mencionados, es disponibilizar conjuntos de datos públicos en portales de sitios web gubernamentales en formato abierto, con el fin de difundir la transparencia ciudadana y reforzar el vínculo entre Estado y ciudadanos. Si bien este paradigma se encuentra establecido en varios países del mundo, aún es necesaria la incorporación de datasets en varios organismos que todavía no han incursionado en dicho tema. Uno de los grandes problemas encontrados en estos portales de datos abiertos, es la inconsistencia de algunos de los datasets, esto se debe a que no todos los conjuntos se encuentran bajo un determinado estándar, sino que los organismos publican en base a guías de sugerencias y/o un criterio predefinido por ellos mismos, es decir, varios organismos podrían tener distintos criterios. Básicamente, no existe un estándar internacional definido obligatorio en el que puedan respaldarse y trabajar en común. Este inconveniente, hace que varios datos presenten un bajo grado de calidad en el dato y luego estos puedan tener inconvenientes al tratar de ser reutilizados para una revisión, análisis o bien algún tipo de interoperabilidad en otras fuentes de datos que alimentan a otros softwares.

Según [OSZ14], quien reflexiona sobre las capacidades que debe tener un Estado, indica que tener capacidades institucionales significa “poseer la condición potencial o demostrada de lograr un objetivo o resultado a partir de la aplicación de determinados recursos y, habitualmente, del exitoso manejo y superación de restricciones, condicionamientos o conflictos originados en el contexto operativo de una institución”. El desafío del Gobierno es brindar a los ciudadanos información de manera eficiente y transparente. En este contexto, los nuevos retos que se plantean pueden ser considerados como una oportunidad para replantear las metodologías de diseñar e implementar políticas públicas e impulsar un Estado con funcionarios colaboradores de la ciudadanía. Todo esto conlleva a la necesidad de nuevas formas de liderazgo colaborativo, y un nuevo paradigma en el tratamiento de la información abierta y pública. Este escenario, demanda nuevas capacidades, tecnologías, equipos y estructuras de trabajo que necesitan promover la innovación como una “nueva actitud” y hacer frente a los nuevos desafíos de la ciudadanía digital.

Es importante identificar que los pilares para la solución de esta problemática son 3 conceptos fundamentales:

- Comprender el contexto de Gobierno Abierto: apoyo del Estado Nacional y Ciudadano;
- Conjunto de datos públicos: organismos estatales deberían ofrecer una mayor cantidad de fuentes de datos de diversos temas gubernamentales con criterios preestablecidos;
- Datos abiertos: mediante la apertura de los diversos datasets en los portales de datos, utilizados como insumo fundamental de información y servicios.



Ramírez Alujas [ALU10] indica que cuando se trata la innovación en el sector público, se debe tener en cuenta el tratamiento de las siguientes cuestiones:

- Nuevas ideas (o conceptos) aplicadas involucra un “hacer”;
- Responder a una necesidad específica o problema;
- Involucrar cambios (deliberados) en el comportamiento de los agentes, las prácticas, procesos y/o productos (bienes y/o servicios públicos);
- Mejorar el desempeño institucional y generar resultados observables;
- Agregan valor público, contribuye al fortalecimiento institucional y mejorar a la calidad de vida de los ciudadanos;
- Mejorar los índices de eficacia, eficiencia y calidad en la prestación de servicios público;

Por lo anteriormente explicado, [ALU10] sostiene algunos problemas que surgen en el Estado como ser: a) “*Procesos mal definidos*”; b) “*Ritmo y la magnitud del cambio estatal*”; c) “*Ausencia de capacidad para el aprendizaje organizacional*”; d) “*La resistencia del público al cambio*”; e) “*Ausencia de recursos*”: falta de apoyo financiero o de carencia de conocimientos para soporte a las iniciativas innovadoras; f) “*Obstáculos técnicos*”: Ausencia de soluciones tecnológicas para el problema en cuestión a analizar. Es decir, es necesario implementar una serie de criterios o aspectos consensuados y sostenibles en el tiempo para mejorar la calidad de los datos públicos abiertos que se ofrecen, lo que brindará mayor poder de análisis de la información dada, como así también, nuevos escenarios sobre posibles mitigaciones de inconvenientes ciudadanos. Esto fomentaría la participación de las personas y/o entidades estatales de todo el mundo en el concepto de transparencia.

A modo de resumen, se puede decir que mantener la calidad y la apertura de los datos públicos, ayuda a los gobiernos y a los diferentes actores de la sociedad civil a tomar mejores decisiones, ya que tienen una visión e información de la realidad. También permite monitorear procesos, construir indicadores y hacer transparente la forma en que se gasta el dinero público [CRO20]. Uno de los puntos a considerar en los portales de datos abiertos es que la disponibilidad de los datos no necesariamente coincide con que tengan calidad, lamentablemente, hoy sigue siendo una dificultad y es un gran desafío para las políticas públicas. El análisis de muchos de los conjuntos de datos públicos representa un problema crucial, ya que está disperso, no estandarizado y en muchos casos desactualizado. Lo mismo puede decirse de la información necesaria para monitorear la transparencia en las compras públicas y la evolución de la actual pandemia de COVID-19 [ZAP20]. El problema en este contexto también tiene que ver con la actualización de la infraestructura técnica para los datos abiertos, ya que las prácticas de gestión de datos son débiles e inconsistentes y cambian con demasiada frecuencia. Los gobiernos deberían trabajar en la transformación técnica y organizacional de los datos abiertos, es decir, invertir en herramientas de hardware y software, como así también en estándares técnicos, capacitación, transformación organizacional y procesos de toma de decisiones para dar un soporte adecuado a la gestión de datos [WEF18]. Como se mencionó anteriormente, aún existen algunas barreras que van desde la calidad insuficiente de los datos publicados hasta la falta de mantenimiento de los portales donde se publican [VEL19]. Varias de estas investigaciones [AOU18], están orientadas a la apertura de un dataset, pero no a un análisis de los problemas que



puedan tener los datos abiertos que están implementados y publicados en los portales gubernamentales.

1.3. Objetivo General

El objetivo principal de la presente tesis es proponer una serie de métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales.

1.4. Objetivos Específicos

Para llevar a cabo dicho objetivo se plantean los siguientes objetivos específicos:

- Estudiar el alcance de los conceptos implicados en este contexto de estudio;
- Analizar los trabajos relacionados en la temática de Gobierno Abierto relativos al tratamiento de datos abiertos y públicos;
- Relevar los trabajos relacionados en cuanto a los aspectos de calidad de datos abiertos y públicos;
- Analizar los criterios de los portales de datos abiertos para efectuar la publicación de éstos;
- Relevar los distintos tipos y formatos disponibles de datos abiertos que existen en los dataset actuales de los sitios más relevantes y gubernamentales de Argentina;
- Analizar las falencias de los datasets disponibles en los sitios de Gobierno;
- Establecer criterios estándar de calidad de datos;
- Proponer métricas críticas y no críticas de calidad de datos abiertos para validar diversos datasets;
- Desarrollar una herramienta de validación de datasets gubernamentales;
- Definir y recolectar la muestra de datasets para ser testeados por la herramienta propuesta como validadora de calidad;
- Definir un análisis cuantitativo en la medición de las métricas propuestas a través de la herramienta desarrollada;
- Analizar los resultados arrojados con la herramienta programada para las métricas de calidad propuestas;
- Validar las métricas sugeridas en base a criterios estándares.



1.5. Hipótesis

La hipótesis de esta tesis es la siguiente:

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 y sencilla.

1.6. Alcance

La idea de Gobierno Abierto no es un tema nuevo, pero involucra diversos conceptos que se encuentran en pleno auge a nivel mundial, tales como los pilares que componen dicha cuestión, siendo: la transparencia, la innovación, la participación ciudadana, rendición de cuentas y la colaboración.

Los gobiernos deben estar abiertos a los ciudadanos, ya que tienen el derecho de acceder a todos los recursos necesarios en cuanto a rendición de cuentas y, además, poder tener la posibilidad de participar en el diseño e implementación de políticas públicas y verificar el estado situación de diversas cuestiones a nivel país.

Esta tesis se enfoca en el tratamiento de los conjuntos de datos públicos gubernamentales en formato abierto de los sitios web estatales, con el fin de validar, fomentar la transparencia pública y empoderamiento ciudadano para participar de asuntos “abiertos al público”, y así obtener un amplio panorama de la situación actual como parte del conocimiento ciudadano. Además, se realizaría un foco especial en la posibilidad de reutilizar dichos datos en otros sitios web y/u organismos, lo que permitirá mejorar la eficiencia de la gestión gubernamental y colaborativa vigente.

1.7. Contribución Científica

Las contribuciones principales de este trabajo de tesis son las siguientes:

- Relevamiento del estado situación actual de los aspectos más relevantes en el tratamiento de datasets públicos abiertos a nivel nacional como internacional.
- Propuesta de un conjunto de métricas críticas y no críticas para analizar la calidad de datos abiertos;
- Desarrollo propio de una herramienta para validar desde varios aspectos estándares, la calidad de los datasets publicados en portales de datos abiertos estatales;
- Detección de falencias en los datasets gubernamentales disponibilizados;



- Contribuir en el análisis, verificación y comprensión del estado actual de los valores que contienen los datasets generados por las entidades gubernamentales más relevantes de la Argentina.
- Aporte sobre las mejoras en la calidad del dato y concientización de su importancia para una correcta divulgación del contenido público tanto a nivel nacional como internacional.

1.8. Enfoque Metodológico

En la Figura 1.3 se muestra el proceso metodológico de investigación empleado para la presente tesis doctoral. Se define un proceso sistemático cualitativo, que implementa una forma evolutiva incremental en cada una de las etapas involucradas.

A continuación, se describen cada una de las etapas del enfoque:

1. **Identificación del Problema:** Definición del problema a tratar, la cual se explicó en la sección 1.2.;
2. **Revisión Teórica:** Búsqueda, selección, recopilación, revisión y análisis del marco teórico y los trabajos relacionados pertinentes al contexto del tema de la tesis;
3. **Recolección de Datos:** Mediante el estudio de la revisión teórica, se definió un alcance de muestra y se extrajeron datasets de los distintos portales de datos abiertos gubernamentales en Argentina;
4. **Clasificación y Análisis de Datos:** Se realiza una segmentación de los datasets de los portales abiertos y se analizan los resultados en aspectos de métricas de calidad de los datos.
5. **Formulación de la Solución:** Se definió el alcance de la propuesta de métricas críticas y no críticas en base al estudio general realizado, identificando niveles de criticidad para éstas.
6. **Aplicación de la Solución:** Se desarrolló una herramienta de software para obtener de forma automática, un análisis cuantitativo para cada una de las métricas de calidad propuestas, a través del relevamiento de datasets extraídos de los portales gubernamentales de datos abiertos;
7. **Validación de la Solución:** Se realizó la validación de las métricas sugeridas en base a los distintos aspectos definidos que debe reunir un conjunto de métricas para la calidad de datasets abiertos. Esta definición parte del estudio realizado sobre la ISO/IEC 25012, dimensiones de calidad, el estándar universal de calidad de datos de dos capas, trabajos de estudio propio en artículos de investigación y la bibliografía relevada en esta tesis.



Figura 1.3 – Enfoque Metodológico de la Tesis Doctoral.

Cabe destacar que la estrategia/metodología formal que permite identificar la investigación es del tipo “investigación-acción”, ya que este tipo de metodología se encarga del estudio de una problemática social específica, siendo en este caso el enfoque de la calidad del contenido de los datasets con información pública del estado nacional gubernamental (Información relevante a obras públicas, salud, educación, tecnologías, etc.). Este tipo de investigación propone una solución a un problema detectado, en este caso, las falencias reveladas en la calidad del contenido de los datos públicos abiertos.

En lo que respecta al planteo de una solución, esta metodología utilizada en la tesis presenta una propuesta de herramienta de desarrollo propio que permite validar los datos abiertos y, además, brindar un estudio detallado sobre lo encontrado en dicha verificación a modo de sugerencia de buenas prácticas.

En la siguiente sección se detalla la organización de la presente tesis y el contenido de cada uno de los capítulos.



1.9. Organización de la Tesis

La presente tesis doctoral cuenta con 7 capítulos de los cuales se complementan con Anexos. Los mismos se detallan a continuación:

- Capítulo 1 - Introducción: Se presenta el contexto introductorio en el que surge esta tesis, se explica de manera breve los conceptos teóricos de Gobierno Abierto, Datos Abiertos y Datos Públicos. Luego, se define el problema, el objetivo principal y los objetivos específicos, posteriormente la hipótesis, el alcance, la contribución científica, el enfoque metodológico utilizado, y finalmente la organización de la estructura de la presente tesis.
- Capítulo 2 - Marco Teórico: Presenta la explicación conceptual de los términos y aspectos que sustentan la temática y las técnicas que serán abordadas a lo largo de la tesis.
- Capítulo 3 - Trabajos relacionados: Se muestra un relevamiento de los trabajos orientados en la experiencia de la temática del contexto teórico de la presente tesis, como así también, aspectos claves de calidad para tener en cuenta, basados en diversos estudios relevados.
- Capítulo 4 - Definición de Métricas para el análisis de Datasets Gubernamentales: Se proponen métricas críticas y no críticas junto con su justificación, importancia y aporte en su utilización, con el fin de establecer un análisis cuantitativo de la validación de aspectos de calidad de los datasets disponibilizados en el ámbito público.
- Capítulo 5 - Implementación: En base a la propuesta de métricas del capítulo anterior se implementa el desarrollo de una herramienta de validación denominada HEVDA, la cual permite automatizar el proceso de validación que se realizaría en forma manual. Para este capítulo se explica cada una de las funcionalidades y cuestiones técnicas de la herramienta programada, como así también se ejemplifica cada métrica propuesta en la aplicación desarrollada. Posteriormente, se lleva a cabo la recolección de datos en donde se cuenta como fue el proceso de selección de los datasets empleados, que luego fueron testeados con ésta. En una sección siguiente, se muestran los resultados arrojados con los estudios realizados sobre el cumplimiento e incumplimiento de dichas métricas de calidad, como así también un breve análisis de los tipos de bloqueantes encontrados en los conjuntos de datos.
- Capítulo 6 - Validación: En este capítulo se presenta la validación de las métricas propuestas en base a los distintos aspectos definidos que debe reunir un conjunto de métricas para la calidad de datasets abiertos. Esta definición parte del estudio realizado sobre la ISO/IEC 25012, dimensiones de calidad, el estándar universal de calidad de datos de dos capas, estudios propios de investigación realizados a entidades organizacionales y la bibliografía relevada en esta tesis.
- Capítulo 7 – Conclusiones y Trabajos Futuros: Se presentan las conclusiones alcanzadas en base a los resultados obtenidos y hallazgos detectados. Además, se indican las líneas futuras de investigación, como así también las publicaciones realizadas en el contexto de la presente tesis.



Los capítulos de este trabajo se complementan con los siguientes Anexos:

- Anexo A – ISO/IEC 25012
- Anexo B – Principios Internacionales de los datos abiertos
- Anexo C – Funciones secundarias de la Herramienta HEVDA
- Anexo D – Portales de datos abiertos gubernamentales
- Anexo E – Detalle de los resultados arrojados con HEVDA

Luego de los siete capítulos y sus anexos, se presentan algunos aspectos, como ser:

- Acrónimos: Se presentan los acrónimos empleados en la tesis.
- Referencias Bibliográficas: Se identifican los libros, revistas, artículos académicos y sitios web gubernamentales utilizados.



Capítulo 2 - Marco Teórico

En este capítulo se presenta la explicación conceptual de los términos sobre los cuales se sustenta la temática y las técnicas que serán abordadas a lo largo de la tesis.

2.1. Innovación Pública

La tecnología avanza velozmente en el estado nacional, y este nuevo concepto, proporciona un rápido acceso a gran parte de la información pública de forma inmediata y en algunos casos, transparente. Sin embargo, es necesaria la mejora de varios procesos existentes en la administración del gobierno público, ya que existen grandes déficits en algunas funciones que están determinadas por modelos administrativos antiguos, y, además, muchos ministerios estatales trabajaban con un paradigma orientado a un régimen 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 contexto reside en la ausencia de información, eficiencia y transparencia, y, también, por la falta de respuestas públicas expeditivas en varias ocasiones. 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” [OSZ14].

Toda esta información disponible lleva a los gobiernos a legislar diferentes leyes que brinden un marco regulatorio sobre la privacidad de los datos a los que se puede tener acceso, como a su vez, al reconocimiento de la disponibilidad de éstos como un derecho para el ciudadano, el cual lo empodera para permitir que los controle, fiscalice y colabore en la toma de decisiones, es decir, forma parte de este nuevo concepto de Gobierno Abierto.

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” [ROD18]. Estos espacios-fronteras permiten la comunicación entre distintos actores con diferentes conocimientos e intereses para lograr un entendimiento común y un aporte ciudadano de forma colaborativa.

“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” [ACE16]. Este concepto se ve reflejado en un cambio cultural, que 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 pública.

Por otra parte, 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” [GOR15].

Según INAP [INA19], básicamente lo que se busca con la innovación pública es:

- Promover el desarrollo de las políticas, instrumentos, capacidades y plataformas necesarias para acelerar los procesos de innovación abierta en el sector público y el crecimiento de un ecosistema de innovación pública y cívica;
- Desarrollar e implementar instrumentos metodológicos para la identificación, formulación y aceleración de proyectos de innovación;
- Realizar conferencias, jornadas de trabajo y aplicar metodologías tendientes a favorecer la circulación de ideas y talentos entre Gobierno Nacional, las distintas jurisdicciones y la sociedad civil.

Uno de los focos centrales del Estado Nacional para llevar a cabo el paradigma de Gobierno Abierto, es la Innovación Pública. Para los ciudadanos involucrados en esta temática se debe crear y construir estructuras organizacionales que promuevan la innovación como una cultura estatal, con el fin de gestionar los nuevos desafíos que presenta una ciudadanía digital.

2.2. Gobierno Abierto

Partiendo del concepto que es necesaria la participación de los ciudadanos en el Gobierno, es que nace el concepto de Gobierno Abierto para la Innovación Pública. Básicamente, la rapidez en el proceso de gestión de los servicios otorgados por el Estado Nacional debe darse a conocer mediante la participación de los ciudadanos en la evaluación y el control de los diversos programas del Estado y de las instituciones públicas, con el fin que el Gobierno mantenga la confianza y un vínculo renovado y activo entre los intereses de los pobladores.



Con este cambio de modelo, el Gobierno Abierto se enfoca en la “construcción de estados transparentes, participativos, que rindan cuentas de manera adecuada e innovadores, poniendo al ciudadano en el centro de la toma de las decisiones públicas, como una forma de fortalecer el Estado democrático. El Gobierno Abierto se basa en cuatro pilares fundamentales: transparencia, rendición de cuentas, participación ciudadana y colaboración e innovación pública y ciudadana” [ARR17]. Este contexto implica la utilización de diversas tecnologías abiertas con el fin de fomentar la Innovación Pública, por ejemplo, “las redes, los sistemas de información, las herramientas inteligentes de búsqueda, el almacenamiento masivo de datos (data-warehousing), Internet, inteligencia artificial, internet de las cosas, big data entre otras, se pueden utilizar para desarrollar sistemas que posibilitan la gestión del conocimiento en las organizaciones públicas, ofreciendo mejores respuestas a las demandas de los ciudadanos, interacción en tiempo real, multicanales de comunicación y distribución de servicios, foros, chats, entre otras; a través de Internet y telefonía móvil” [DAR18].

Por otra parte, se sostiene que este término “es una doctrina política que surge a partir de la adopción de la filosofía del movimiento del software libre a los principios de la democracia. El Gobierno Abierto tiene como objetivo que la ciudadanía colabore en la creación y mejora de servicios públicos y en el robustecimiento de la transparencia y la rendición de cuentas” [SOS21]. El Estado Nacional debe incorporar las técnicas, TIC y los conocimientos necesarios para gestionar dicha cantidad de datos con un diseño tecnológico y por sobre todo enfocado en el ciudadano promedio para fomentar la inclusión social.

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” [QUI16].

2.3. Datos Públicos

La información publicada de manera transparente y explícita a todos los ciudadanos es previamente evaluada, procesada, depurada y provista por algún grupo u organización estatal. “El dato público es todo aquel que puede ser accedido o consultado por cualquier persona u organismo de forma libre, aunque no necesariamente sean datos digitalizados, en Argentina, para que un dato pueda ser público debe hallarse dentro del marco regulatorio de las siguientes leyes” [SEC21]:

- Ley de Protección de Datos Personales 23.526/00;
- Ley del Sistema Estadístico 17.622/68;
- Ley de Procedimiento Fiscal 11.683/32;
- Ley de Propiedad Intelectual 11.723/33.

Manfredi manifiesta que “en el plano constitucional, el acceso a la información es el desarrollo natural del ejercicio de los derechos individuales de la esfera pública. El acceso actúa como precondition al ejercicio de otros derechos y como fuente de ciudadanía política. Influye sobre el criterio y sobre la capacidad de ejercer los derechos humanos” [MAN17], y esto van en consonancia con la Ley 104 [LEG16], que hace hincapié sobre el derecho al acceso a la información pública.



El programa de datos abiertos en ciencia y tecnología de la República Argentina, a través de la Resolución 640-E/2017 [BOL17], 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.

Cada vez son más los portales estatales que proporcionan datos públicos ya sea para usuarios finales como también para alimentar otros sistemas. Para cada conjunto de datos públicos es necesario determinar su valor, su calidad y quién es el responsable de su mantenimiento. “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” [MON17].

La Agenda 2030 [AGE15], [AGE21] para el Desarrollo Sostenible es el documento A/RES/69-315 [NAC15] 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, que 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 e innovación con la ciudadanía para brindar ayuda a estas necesidades. Por ello, para realizar un proceso de apertura serio y útil, es vital que los datos sean claros, completos, confiables y precisos.

2.3.1. Acceso a la Información Pública

El enfoque del nuevo paradigma permite que un gobierno transparente, ponga a disposición de los ciudadanos, información sobre sus acciones y planes de forma inmediata, fácil y gratuita. “Al ampliar el acceso a la información pública se fortalece la rendición de cuentas y se enriquece el debate público, a la vez que se crean nuevas oportunidades para generar valor agregado” [KIT17].

La 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 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 [LEY16]. Para el caso de la República Argentina, la cooperación técnica entre las distintas



jurisdicciones y la concentración de políticas en lo que respecta a la transparencia y el acceso a la información pública, depende del Consejo Federal para la transparencia [CON16]. Este Consejo Federal para la Transparencia fue creado por el artículo 29 de la Ley 27.275 [LEY16] de acceso a la información pública y está constituido por un representante de cada una de las provincias y un representante de la Ciudad Autónoma de Buenos Aires.

La transparencia puede ser de dos tipos: pasiva o activa, esto depende si la información es solicitada a demanda por un ciudadano u organismo (Pasiva), o bien, si el Estado la disponibiliza públicamente (Activa).

2.3.1.1. Transparencia Activa

La transparencia activa es un concepto en el que los distintos organismos públicos, deben dar acceso a la información y tienen la obligación de facilitar todos estos datos a través de su sitio web institucional, es decir, publicar y difundir periódicamente información relevante en un formato accesible y abierto.

2.3.1.2. Transparencia Pasiva

La transparencia pasiva se asocia con el derecho del acceso a la información por parte de los ciudadanos. Esto implica la garantía del derecho de acceso a la información que tenemos todas las personas, como lo estipula la Declaración Universal de los Derechos del Humanos (art. 19) [DEC21] y el Pacto Internacional de Derechos Civiles y Políticos (art. 19,2) [PAC21]. Básicamente, propone un medio institucional para que los ciudadanos puedan solicitar la información que produce el Estado. En Argentina está garantizado por la Ley 27.275 de Acceso a la Información Pública [LEY16].

Este trabajo tiene un enfoque especial en la calidad de los datos abiertos que son brindados a través de los portales abiertos de distintos organismos estatales es por lo que, en la siguiente sección se analizan las características de los datos abiertos y luego, distintos aspectos para tener en cuenta en la calidad de estos conjuntos de datos públicos.

2.4. Datos Abiertos

El Ministerio de Ciencia, Tecnología e Innovación de la República Argentina [MIN20], señala que el formato abierto, es el formato de archivo no propietario, cuya especificación debe estar documentada públicamente, es de libre conocimiento e implementación y libre de patentes o de cualquier otra restricción legal o económica para su uso. Para ello, es necesaria una licencia abierta, que es un acuerdo de provisión de datos para que cualquier persona los utilice, los reutilice y los distribuya, estando sujeto a las condiciones de dicha licencia.

“Los datos que se pueden reutilizar y redistribuir sin ninguna restricción se denominan datos abiertos” [OVI13]. Estos datos deben estar en formatos digitales, con un modelo estándar abierto. Por otra parte, es importante destacar que, “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” [GAR11]. “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. Básicamente, se trata de poner a disposición de la sociedad, de manera libre, una 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” [NAS17]. 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. Debido a todo lo anteriormente explicado, es fundamental que personal dedicado trabaje en el tratamiento de las fuentes de datos abiertos, para que los organismos estatales y ciudadanos de la República Argentina, tengan un mejor conocimiento sobre un determinado tema público, como ser: economía, transporte, etc.

A la hora de hablar de datos abiertos se hace referencia a grandes volúmenes de información clasificados y organizados en datasets específicos, Venegas manifiesta que “los repositorios de datos abiertos son grandes bancos de información que se encuentran disponibles a todo el mundo, albergando datos de todos los temas y formatos” [VEN17].

Mediante el acceso a los portales de datos abiertos, se pueden extraer datasets de diferentes categorías de información. Algunos portales gubernamentales presentan APIs para obtener principalmente información en tiempo real, de transporte, tránsito, entre otros. Otros tipos de archivos públicos utilizados para la apertura suelen entregarse en formatos estandarizados como, por ejemplo: CSV, XLS, XLSX, XML, entre otros. La mayor diversidad de formatos ofrecidos facilitará que un tercero pueda elegir el formato conveniente para visualizar los datos o utilizarlos como entrada de otra herramienta o sistema. Fundamentalmente, los datasets publicados en portales de datos abiertos brindan información acerca de una temática específica para retroalimentar a otros sistemas de información, por lo que, su publicación está sujeta a las leyes vigentes de cada región. Por ejemplo, para la Ciudad de Buenos Aires, la información proporcionada a los ciudadanos se encuentra regulada bajo el Decreto 156/2012 [BOL12].

Magallón afirma que “la cultura de datos abiertos trata de obtener un valor añadido de la información. A diferencia de lo ocurrido hasta ahora esta información no genera sólo su valor por estar reservada a unos pocos, sino que lo hace por su disponibilidad para ser interpretada y traducida por cualquier actor interesado en trabajar con ella” [MAG17], dicho en otras palabras, un dato abierto es aquel que puede ser accedido, pero lleva un formato que permite la interoperabilidad con otros softwares. Se puede decir que, “los datos abiertos son el proceso que pone al alcance de la sociedad los datos públicos de los que dispone la Administración Pública, en formatos digitales, estandarizados y abiertos” [GAR11].

2.4.1. Niveles de Apertura de Datos

Existen distintos parámetros para definir el nivel de apertura de los datos. Tim Berners-Lee [OPD12], iniciador de los Datos Enlazados (Linked Data), sugirió un esquema de clasificación a partir de 5 estrellas en la que cada una representa un escalón de desarrollo en la apertura de los datos, del más simple al más complejo. En la Figura 2.1, se muestra el esquema propuesto por él.

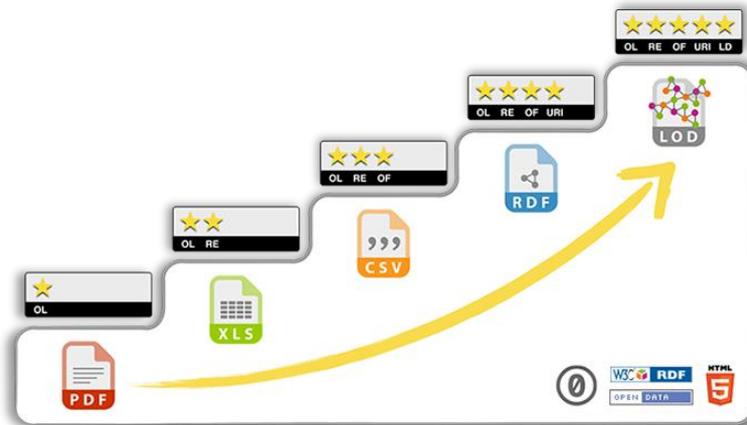


Figura 2.1 –Esquema de evolución por niveles de los formatos abiertos para las 5 estrellas.

Cada categoría está representada con una cantidad de estrellas:

- Una estrella: implica que el contenido web esté disponible (por ejemplo: archivos .pdf o una imagen), pero no permiten la reutilización;
- 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;
- Tres estrellas: están disponibles en un formato abierto no propietario (sin necesidad de pagar por un software) y estructurado (por ejemplo: CSV);
- 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 otra para los propios datos. Esta combinación, permite formar un esquema global;
- 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.

2.4.2. Apertura de datos en el mundo

Hoy es posible acceder a información sobre diversos temas y categorías de instituciones y organismos públicos. “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 (AGA /Open Government Partnership” [CRI19]. El Open Government Partnership, OGP [OGP21] 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 fechas determinadas. En 2011, los líderes gubernamentales y los defensores de la sociedad civil se unieron para crear una asociación única, para promover una gobernabilidad responsable e inclusiva. Según el sitio web oficial de OGP, existen setenta y ocho 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. Esta alianza reúne las diferentes iniciativas de varios gobiernos alrededor del mundo. Su misión es introducir a los gobiernos en la implementación de las mejores prácticas para llevar a cabo un Gobierno Abierto, tanto en sus compromisos de reforma, como en los procedimientos de gestión, habilidades y tecnologías a utilizar. Esto incluye organizaciones no gubernamentales, así como empresas privadas con experiencia [NAV20]. Según el Grupo de Investigación en Gobierno, Administración y Políticas Públicas, GIGAPP [GIG21], una parte esencial del Gobierno Abierto consiste en aplicar el término de transparencia, que brinda información sobre lo que está haciendo el gobierno, cuáles son sus planes de acción y datos fuentes, y lo que se puede responsabilizar ante la sociedad. Esto fomenta y promueve la rendición de cuentas de la administración a los ciudadanos y permite un control social adecuado. Por ello, los gobiernos definen algunas herramientas de política pública para el gobierno abierto, de acuerdo con el principio de transparencia.

Por otro lado, cada vez son más los países que se unen en este nuevo paradigma de diseñar políticas e iniciativas de datos abiertos, en materia de esto el Open Data Impact Map, OD4D [ODI21], es un proyecto de la Red de Datos Abiertos para el Desarrollo [ODD21], 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. A continuación, se muestra la Figura 2.2, con un mapa de 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.

La fundación Open Knowledge International [PKI21] 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.

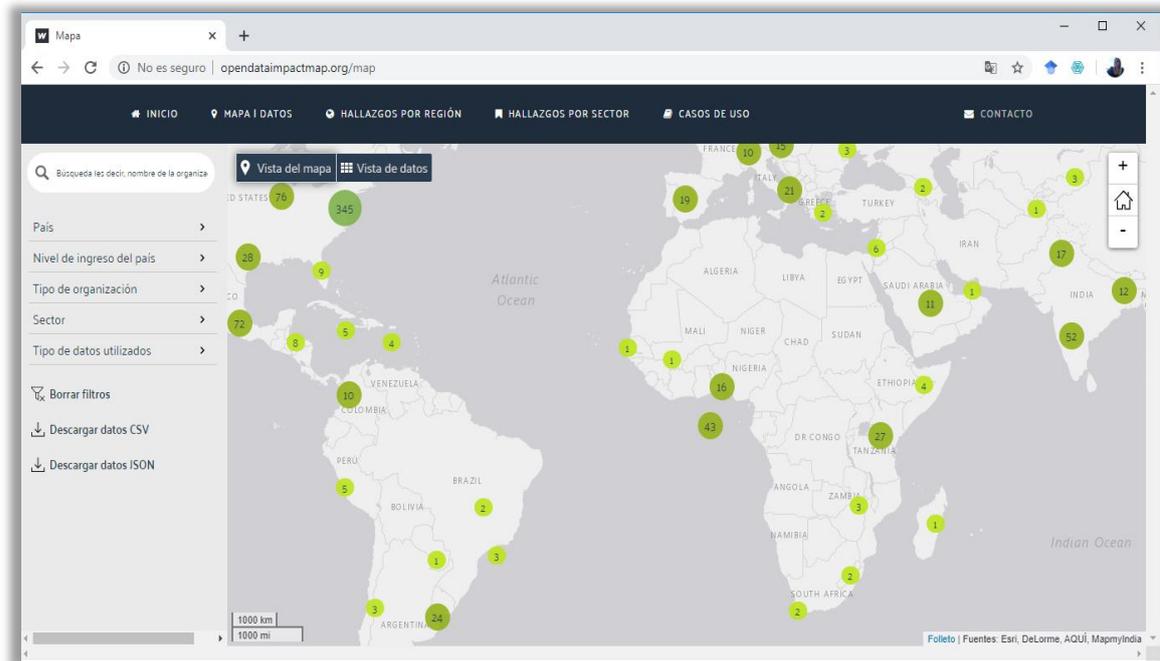


Figura 2.2 – Mapa de impacto de datos abiertos.

Por otra parte, existe un sitio web llamado barómetro de datos abiertos a nivel mundial [ODB21], 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 [ODC21].

El barómetro de datos abiertos [ODB21] 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 2.3, se muestra el mapa mundial con los 30 gobiernos, discriminados por color en base al puntaje que posee cada país. Los tonos van desde el rojo (menor puntaje) hasta el color verde (mayor puntaje). Según el sitio, Argentina posee 47 puntos.

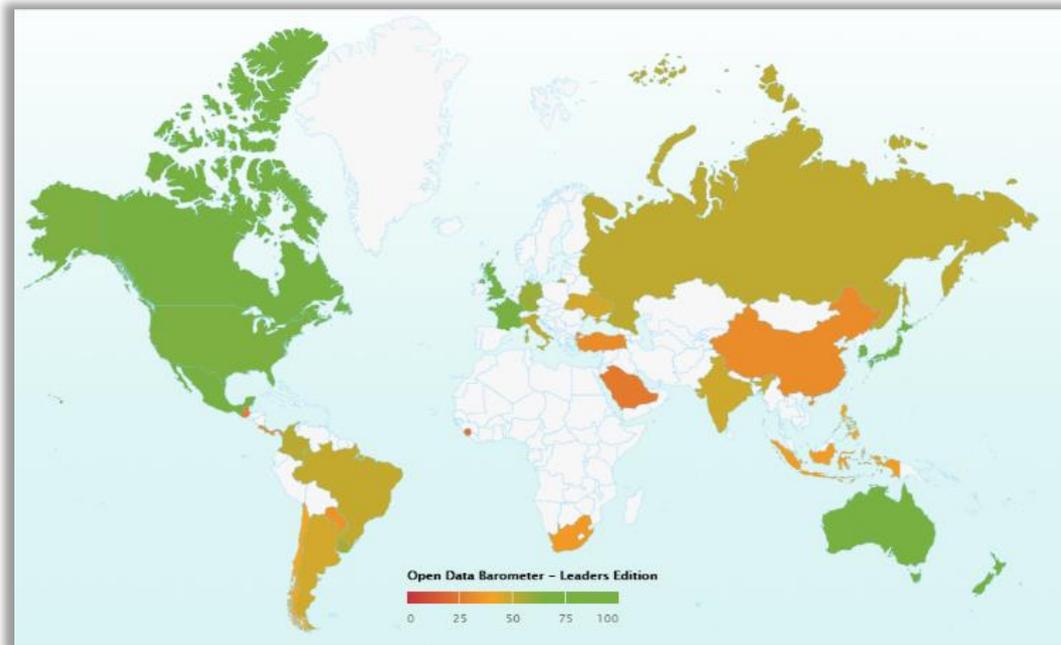


Figura 2.3 – Calificación de los gobiernos en el sitio del barómetro de datos abiertos.

En el caso de las ciudades, los datos abiertos establecen una gran oportunidad para mejorar la eficiencia, la productividad y la creatividad, tanto de los sectores de organismos estatales, como de los privados que sean capaces de explotarlos [TOM19]. Desde aspectos de normativas, la Asociación Española de Normalización y Certificación (AENOR) [AEN21], ha publicado la primera norma para las Smart Cities, la UNE 178301 sobre Datos Abiertos [UNE21], que establece los requisitos para la reutilización de datos elaborados o custodiados por el sector público. Esta norma establece la forma de evaluar la publicación de datos abiertos, por lo que determina métricas y un indicador que permiten medir el grado de madurez de la apertura de datos del sector público de forma que se facilite su reutilización.

2.4.3. Apertura de datos en América Latina y el Caribe

América Latina y el Caribe es una región altamente comprometida con la agenda de datos abiertos. Varios de los países de la región cuentan con una regulación adecuada de acceso a la información pública y han implementado portales de datos abiertos y están elaborando legislaciones sobre protección de datos personales [MUE18]. Existen iniciativas, como la Iniciativa Latinoamericana de Datos Abiertos (ILDA), que busca estandarizar la recolección y publicación de estos datos públicos y promover su reutilización. Uno de los informes de ILDA muestra el análisis del Barómetro Regional de Datos Abiertos para América Latina y el Caribe 2020 [ILD20] que indica que, a pesar de la variabilidad en las políticas públicas de datos abiertos, las comunidades de usuarios de datos se están volviendo más activas y diversas. Hay proyectos más ambiciosos basados en datos abiertos que buscan generar impacto social. Por lo tanto, las comunidades técnicas también están involucradas en temas relacionados con la calidad de los datos públicos, por ejemplo, con los datos de salud pública en el contexto de COVID-19. Este informe abarcó 24 países, 16 de América Latina (Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú y Uruguay) y 8 del



Caribe (Bahamas, Belice, Guyana, Haití, Jamaica, República Dominicana, Santa Lucía y Trinidad y Tobago). Esta medición también clasifica a los gobiernos de estos países según a) Preparación para iniciativas de datos abiertos; b) La implementación de programas; y c) El impacto que están teniendo los datos abiertos.

Según estudios [BID20], los países de la región deberían hacer mayores esfuerzos para mejorar la participación de los actores en la calidad e integridad de los datos, es decir, deberían compartirse más conjuntos de datos de alto valor y mejorar su accesibilidad. Los organismos públicos deberían invertir en la mejora de sus portales de datos y sistemas estadísticos para aumentar la disponibilidad y accesibilidad de sus conjuntos de datos. Según un informe [ZAP20] de CAF, (Corporación Andina de Fomento, Banco de Desarrollo de América Latina) [CAF21], la cooperación internacional y los bancos de desarrollo han sido fundamentales en el avance de esta agenda en la región, fortaleciendo las capacidades del sector público en muchos casos. CAF ha realizado acciones de apoyo a políticas de datos abiertos, por ejemplo, en la Ciudad de México para el uso de datos para modelos predictivos ante la pandemia Covid-19; en Ecuador para la implementación del Programa Interamericano Anticorrupción de Datos Abiertos; en Colombia se trabajó en un modelo de inteligencia de datos; y en Buenos Aires, donde avanza hacia una nueva etapa en su política y estrategia de reutilización de datos abiertos.

2.4.4. Apertura de datos abiertos en Argentina

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 más información, se puede consultar la fuente [OGP12], con todo lo que respecta a su compromiso con el Gobierno Abierto.

En la Ciudad de Buenos Aires, Argentina existe la Ley 104 de acceso a la información pública, en la cual se expone el derecho del ciudadano a obtener información pública [LCA16], esta Ley brinda un reconocimiento al acceso de la información como un derecho humano y, además, formaliza la apertura de datos al ciudadano. En el año 2013 la ciudad autónoma de Buenos Aires publicó el Decreto 478 [DEC13] 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. A nivel nacional, la legislación, a través del Decreto 117/2016 [DEC16], estableció la apertura de los datos públicos del Poder Ejecutivo Nacional y su publicación en el portal de datos: datos.gob.ar.

Argentina es uno de los países que incentiva la participación ciudadana, además de ofrecer eventos y charlas abiertas a ciudadanos, empresas privadas y estatales. Una de las últimas reuniones fue la realizada por la Ciudad Autónoma de Buenos Aires, sobre el IV Plan de Acción de Gobierno Abierto 2019-2022 [CUA21] que consta de 18 compromisos que buscan desarrollar políticas transformadoras para resolver problemas específicos de la agenda pública. Además de los compromisos asumidos por el país, se asignaron las áreas responsables de llevar a cabo este trabajo. Los participantes de este evento fueron: la Dirección Nacional de Gobierno Abierto, por parte del Jefe de Gabinete de Ministros [GOB21], la Auditoría General (AGN) [AGN21], con la participación especial del responsable de la Área de Participación Ciudadana, quienes agregaron nuevos enfoques en el área de auditoría del país, con el fin de apoyar y promover la transparencia en los datos de los diferentes organismos públicos, y el alcance de la iniciativa de Gobierno Abierto hacia ellos. Por otro lado, otro de los actores en la sensibilización de este nuevo paradigma es la



Agencia de Acceso a la Información Pública [AAI21], esta Dirección promueve algunas variables para la publicación de información pública, tales como: Declaraciones juradas, Compras, presupuestos, entre otras. En el que el objetivo principal es proporcionar un canal de comunicación con el público para asesorar en las solicitudes de información pública y colaborar en la dirección de la solicitud y afinar la búsqueda (plataforma tecnológica, marco legal, etc.). Otro de los actores involucrados en este tema es el Consejo Nacional de la Magistratura [CMG21], a través de la Unidad de Consejo Abierto y Participación Ciudadana [CMF21], quienes aseguran que el estallido de la Pandemia se ha convertido en una nueva oportunidad para avanzar en diversos diseños de políticas públicas relacionados con el concepto de Gobierno Abierto, por lo que este genera un cambio de comportamiento en los distintos organismos públicos. Otra de las entidades involucradas es la Honorable Cámara de Diputados de la Nación Argentina [HCD21], como otro núcleo central que actualmente promueve la importancia de la información pública y la transparencia en el país. Finalmente participó la Fundación Huésped de Argentina [FHP21] en representación de la Sociedad Civil, y especialmente en materia de conocimiento público sobre temas relacionados con el coronavirus, la Fundación es una organización argentina con alcance regional que viene trabajando desde 1989 en áreas de salud pública desde una perspectiva de derechos humanos enfocadas en VIH ((virus de inmunodeficiencia humana), otras enfermedades transmisibles y salud sexual y reproductiva. Esto es parte de la mesa nacional de Gobierno Abierto, ya que cada país debe tener un vínculo entre Gobierno y Sociedad. Estos afirman que no todas las organizaciones tienen los mismos intereses o recursos para poder participar en los temas de Gobierno Abierto, por lo que es necesario que las organizaciones con menos recursos se puedan articular con las de más recursos para poder empoderarse en este contexto.

El portal oficial del gobierno de datos abiertos en Argentina [AUD21] presenta aproximadamente 1.050 conjuntos de datos. Se presentan guías sobre varios temas, tales como: recomendaciones para publicar datos en formatos abiertos [GUP21], cómo estructurar una tabla, nombrar archivos y columnas, utilizar estándares básicos y trabajar con hojas de cálculo. Por otro lado, se puede acceder a guías de buenas prácticas para el uso y publicación de metadatos [GUU21], y guías de apertura de datos en organismos de la Administración Pública Nacional, así como en Gobiernos Subnacionales. Por otro lado, se presentan documentos y algunos recursos para promover la apertura y el consumo de datos públicos [GUH21]. Sobre la calidad de los conjuntos de datos, se presentan buenas prácticas para mejorar la calidad de los conjuntos de datos, por ejemplo, se indica que deben estar en formatos abiertos habilitados, por ejemplo: CSV, XML, JSON, entre otros. Para los desarrolladores de software, las rutinas de transformación y limpieza de datos que siguen las buenas prácticas se pueden implementar a través de una herramienta llamada Data Cleaner [DCL21], que se encuentra en la etapa de prueba. Esto permite el uso de plantillas de secuencias de comandos para limpiar o reemplazar caracteres, indicar saltos de línea, eliminar columnas y filas, normalizar tipos de fechas y separar campos específicos de un conjunto de datos. Si bien es una herramienta interesante, no proporciona un análisis detallado de la calidad del conjunto de datos. Uno de los aspectos agregados por el Gobierno argentino es la estandarización en algunos campos determinados como interoperables [GUI21], para que los datos puedan ser utilizados e identificados a nivel de software. Esto solo aplica a datos del tipo de personas físicas y jurídicas que siguen la norma, así como datos geográficos (provincias, departamentos, municipios, localidades, calles o coordenadas) en los que se puede utilizar una API del servicio de Normalización de Datos Geográficos [API21].



2.4.4.1. Relevamiento de Portales abiertos gubernamentales provinciales argentinos

Argentina se divide en 23 provincias, para cada una de éstas se realizó un relevamiento de su portal oficial a nivel Nacional de cada provincia, teniendo en cuenta los aspectos de Gobierno Abierto y Transparencia al ciudadano. A continuación, se indica lo relevado en aspectos de datos abiertos disponibilizados a los ciudadanos y organizaciones tanto públicas como privadas:

1. Buenos Aires

La provincia de Buenos Aires presenta en su sitio oficial de Datos Abiertos [PBA21], llamado “Datos Abiertos PBA”, unos 67 conjuntos de datos (datasets), los cuales se encuentran enfocados en 12 categorías (“Administración Pública”, “Agricultura y Ganadería”, “Cultura”, “Demografía y Territorio”, “Economía y Finanzas”, “Educación”, “Género”, “Industria”, “Medio Ambiente”, “Movilidad y Transporte”, “Salud” y “Seguridad y Justicia”). 17 organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos fue desarrollado con Andino [AND21] con CKAN [CKA21], Versión: 2.5.6, y cuenta con filtros de búsqueda por nombre de dataset, datos claves de etiquetas y filtrado de organizaciones gubernamentales. Algunos de los tipos de formatos abiertos son: XML, CSV y XLS. Si bien el formato PDF no se considera abierto para interoperabilidad para software, también existen algunos archivos de este tipo.

2. Catamarca

La provincia de Catamarca presenta en su sitio oficial de Datos Abiertos [GCA21], llamado “Datos Catamarca” unos 4 conjuntos de datos (datasets), los cuales se encuentran enfocados en una categoría, siendo ésta “Economía y Finanzas”. 2 organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos está desarrollado con Andino [AND21] con CKAN [CKA21], Versión: 2.6.3, y cuenta con filtros de búsqueda por nombre de dataset, datos claves de etiquetas y filtrado de organizaciones gubernamentales. Los tipos de formatos abiertos son: CSV y XLS.

3. Chaco

La provincia de Chaco presenta en su sitio oficial de Datos Abiertos [GCH21], llamado “Gobierno Abierto Provincia de Chaco, Datos Abiertos”, el cual actualmente se encuentra en reparación, ya que se visualiza una leyenda que indica que el sitio se encuentra en construcción. Si bien no posee datos abiertos disponibilizados, esta provincia presenta un enfoque de información sobre el contexto de Gobierno Abierto [PCH21] y su importancia para la colaboración en la transparencia y rendición de cuentas. En dicho sitio web [PCH21], se presentan estadísticas generales de la provincia en aspectos de: presupuesto vigente, obras en construcción, economía, transporte, temas de la sociedad en general (trabajo, salud, etc.), como así también, un espacio para solicitar información pública, datos relevantes para conocer y realizar un seguimiento en cuestiones de rendiciones de cuentas gubernamentales y por último, una sección de participación ciudadana que incluyen guías de trámites y seguimiento de los mismos.

4. Chubut

La provincia de Chubut presenta en su sitio oficial de Datos Abiertos [CHU21], llamado “Datos Abiertos Gobierno de Chubut” unos 18 conjuntos de datos (datasets), los cuales se encuentran enfocados en 12 categorías (“Gobierno y Sector Público”, “Agricultura, pesca, silvicultura y alimentación”, “Ciencia y Tecnología”, “Economía y Finanzas”, “Educación, Cultura y Deportes”, “Energía”, “Justicia, sistema judicial y seguridad pública”, “Medio Ambiente”, “Población y



Sociedad”, “Salud”, “Regiones y Ciudades” y “Transporte”). 9 organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos fue desarrollado con CKAN [CKA21], y cuenta con filtros de búsqueda por nombre de dataset, datos claves de etiquetas, tipo de formatos abiertos, licencias y filtrado de organizaciones gubernamentales. Algunos de los tipos de formatos abiertos son: GeoJSON y CSV. Finalmente, en el sitio web oficial, se presenta una breve sección en la que se explica a grandes rasgos, sobre la importancia de los datos abiertos y el contexto de Gobierno Abierto. Cabe aclarar que existe otro sitio web dedicado especialmente a la concientización de Gobierno Abierto en la provincia de Chubut [GAC21].

5. Córdoba

La provincia de Córdoba presenta en su sitio oficial de Datos Abiertos [GCO21], llamado “Datos Abiertos” unos 120 conjuntos de datos (datasets), los cuales se encuentran enfocados en 4 ejes de gestión (“Crecimiento Económico Sostenible”, “Justicia Social”, “Fortalecimiento de las Instituciones” y “Prosperidad”). Varias reparticiones gubernamentales son las que participan y colaboran con sus datos. Este portal de datos abiertos fue desarrollado por la propia provincia, y cuenta con filtros de búsqueda por Grupos de ODS (Objetivos de Desarrollo Sostenible) [ODS21], Reparticiones, ODS, Ejes de Gestión, Etiquetas, Tipos de formatos y Licencias. Algunos de los tipos de formatos de archivo que utilizan son: SVG, XLSX, CSV y PDF.

Por otro lado, se presenta una sección aparte, destinada a la obtención de datos abiertos y visualizaciones de más datos [COD21], la cual concientiza que la transparencia en la gestión pública es fundamental para lograr una democracia participativa, y además, permite encontrar datos públicos del Poder Ejecutivo de Córdoba para usar y compartir (Presupuestos, Compras Públicas, Resultados electorales, Autoridades Provinciales, Declaraciones Juradas, Portal de Transparencia, caja de Jubilaciones, Estadísticas varias, Lugares Públicos, entre otros). Cabe aclarar que este sitio no tiene una estructura de portal de datos abiertos, sino que son accesos a varios archivos en formato PDF y XLS.

6. Corrientes

La provincia de Corrientes no presenta un sitio oficial de Datos Abiertos en su Portal del Gobierno de Corrientes [COR21]. Cabe aclarar que esta provincia si bien no tiene formalmente un enfoque en Gobierno Abierto y datos abiertos, existe un portal de Datos Abiertos por parte de la Municipalidad de Corrientes [MCO21], llamado “Municipalidad de Corrientes | Datasets” que tiene unos 48 conjuntos de datos (datasets), desarrollado con Andino [AND21] con CKAN [CKA21], Versión: 2.5.6.

7. Entre Ríos

La provincia de Entre Ríos no presenta un sitio oficial de Datos Abiertos en su Portal del Gobierno de Entre Ríos, existe una sección de Gobierno Abierto [GER21], la cual contiene información sobre Transparencia, Participación y Gobierno Digital. Cabe aclarar que esta provincia posee datos públicos que se encuentran disponibilizados a modo de enfoque de Transparencia, pero no así abierta [GTR21]. Esta provincia, si bien no tiene formalmente un portal de datos abiertos oficial, existe un sitio web que tiene Datos Abiertos por parte de la Municipalidad de Gualeguaychú [GGU21], llamado “Gualeguaychú, la ciudad de todos | Municipalidad de Gualeguaychú estadísticas” que tiene unos 246 conjuntos de datos (datasets), gestionado con CKAN [CKA21].

8. Formosa

La provincia de Formosa presenta en su sitio oficial de Datos Abiertos [FOR21], llamado “Datos Abiertos” unos 22 conjuntos de datos (datasets), los cuales se encuentran enfocados en 6 Áreas (“Agenda de Eventos”, “Agenda de Gobierno”, “Capacitaciones”, “Noticias”, “Servicios”,



“Trámites”). Este portal de datos abiertos es de desarrollo propio de la provincia, es sencillo y no cuenta con filtros de búsqueda. Algunos de los tipos de formatos de archivos que utilizan son: XML y RSS, algunos de estos archivos al realizar clic no se encuentran disponibilizados.

9. Jujuy

La provincia de Jujuy presenta en su sitio oficial de Datos Abiertos [JUJ21], llamado “Datos Abiertos”, lamentablemente no se visualiza ningún datasets, ya que se estima que el sitio web está en construcción. Cabe aclarar que, si bien no existen datasets en el sitio web [JUJ21], se tiene opciones enfocadas a Gobierno Abierto y a la concientización de Datos Abiertos, Transparencia, participación y acceso a datos públicos varios (en formato PDF).

10. La Pampa

La provincia de La Pampa no presenta en su sitio oficial de gubernamental [PAM21], un Portal de Datos Abiertos. Cabe aclarar que no se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos.

11. La Rioja

La provincia de La Rioja presenta en su sitio oficial de Datos Abiertos [RIO21], llamado “Dato Abierto | Participación”, unos 403 conjuntos de datos (datasets), los cuales se encuentran enfocados en 18 categorías (“Ciencia y Tecnología”, “Comercio”, “Demografía”, “Economía”, “Educación”, “Empleo”, “Energía”, “Hacienda”, “Industria”, “Legislación y Justicia”, “Medio Ambiente”, “Medio rural”, “Salud”, “Sector Público”, “Sociedad y bienestar”, “Turismo”, “Urbanismo e infraestructura” y “Vivienda”). Varias organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos es gestionado con CKAN [CKA21], y cuenta con filtros de búsqueda por nombre de dataset, Consejerías, Temas, Formatos, Etiquetas, Tipo de Datos, Frecuencia y Licencia. Algunos de los tipos de formatos abiertos son: JSON, API, XML, CSV y XLS.

12. Mendoza

La provincia de Mendoza presenta en su sitio oficial de Datos Abiertos [MEN21], llamado “Datos Abiertos Mendoza”, unos 161 conjuntos de datos (datasets), los cuales se encuentran enfocados en 11 categorías (“Agroganadería”, “Pesca Y Forestación”, “Economía Y Finanzas”, “Educación, Cultura Y Deportes”, “Elecciones”, “Energía”, “Gobierno Y Sector Público”, “Justicia, Seguridad Y Legales”, “Medio Ambiente”, “Municipios”, “Población Y Sociedad”, “Recursos Hídricos”, “Regiones Y Ciudades”, “Salud” y “Turismo”). 24 organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos está desarrollado con Andino [AND21] y gestionado con CKAN [CKA21], Versión: 2.6.0, y cuenta con filtros de búsqueda por nombre de dataset, Temas, Organizaciones y Etiquetas. Algunos de los tipos de formatos que tienen los archivo son: CSV, ZIP, PDF y Docx.rtf (Otro).

13. Misiones

La provincia de Misiones presenta en su sitio oficial de Datos Abiertos [MIS21], llamado “Datos Abiertos” unos 13 conjuntos de datos (datasets), los cuales se encuentran enfocados en 8 organizaciones que son las que participan y colaboran con sus datos. Este portal de datos abiertos se encuentra gestionado con CKAN [CKA21], y cuenta con filtros de búsqueda por nombre de dataset, Organizaciones, Grupos, Etiquetas, Tipos de Formatos de archivo y Licencias. Algunos de los tipos de formatos abiertos que tiene en su sitio son: CSV, GeoJSON, ZIP y OWL.



14. Neuquén

La provincia de Neuquén presenta en su sitio oficial de Datos Abiertos [NEU21], llamado “Neuquén Open data”, con unos 19 conjuntos de datos (datasets), los cuales se encuentran enfocados en organizaciones que son las que participan y colaboran con sus datos. Este portal de datos abiertos se encuentra gestionado con CKAN [CKA21], y cuenta con filtros de búsqueda por nombre de dataset, Organizaciones, Grupos, Etiquetas, Tipos de Formatos de archivo y Licencias. Algunos de los tipos de formatos abiertos que tiene en su sitio son: XLS, CSV, PDF, JPG, DOC, XML, PPT, y JSON. Cabe aclarar que no todos los tipos de formatos de archivos con lo que se disponibilizan son del tipo datos abierto, por ejemplo: el tipo de formato JPG, DOC y PDF.

Por otra parte, en el mismo sitio web [NEU21], se presenta una sección con información sobre la concientización de los datos abiertos y su importancia en Gobierno Abierto, como así también la explicación del porque se utiliza CKAN [CKA21].

15. Río Negro

La provincia de Río Negro no presenta en su sitio oficial de gubernamental [RNE21], un Portal de Datos Abiertos. Cabe aclarar que no se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos, pero si se disponibiliza algunos aspectos de Transparencia, como ser: Boletín Oficial, Compras y Licitaciones, Consulta de expedientes, y Consulta de pago a proveedores.

16. Salta

La provincia de Salta no presenta en su sitio oficial de gubernamental [GSA21], un Portal de Datos Abiertos. Cabe aclarar que no se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos, pero si se disponibiliza algunos aspectos de Transparencia enfocados al ciudadano, como ser: Compras y Contrataciones, Obras Públicas, Escala Salarial, Cronograma de Pagos, entre otros. Cabe aclarar que estos son datos públicos, no así datos abiertos.

17. San Juan

La provincia de San Juan presenta en su sitio oficial de Datos Abiertos [SAJ21], llamado “Datos Abiertos San Juan” unos 32 conjuntos de datos (datasets), los cuales se encuentran enfocados en 7 organizaciones son las que participan y colaboran con sus datos (“Minería”, “Hacienda y Finanzas”, “Gobierno”, “Salud”, “Infraestructura”, “Medio Ambiente”, “Turismo y Cultura”). Este portal de datos abiertos es gestionado con CKAN [CKA21], y cuenta con filtros de búsqueda por nombre de dataset, Organizaciones, datos claves de etiquetas, Formatos y Licencias. Algunos de los tipos de formatos de archivos son: KMZ, XLS, PDF, JSON, XML y CSV.

18. San Luis

La provincia de San Luis no presenta en su sitio oficial de gubernamental [SAT21], un Portal de Datos Abiertos. Cabe aclarar que no se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos, pero si se disponibilizan algunos aspectos de Transparencia enfocados al ciudadano, como ser: “Organigrama”, “Solicitudes y Denuncias” y “Estadísticas y Censos”. Al igual que otras provincias, cabe aclarar que estos son datos públicos, no así datos abiertos.

19. Santa Cruz

La provincia de Santa Cruz no presenta en su sitio oficial de gubernamental [SAP21], un Portal de Datos Abiertos. Cabe aclarar que se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos [SCN21], si bien no se encuentra en formato abierto, pero se



observan algunos aspectos de Presupuestos, Contaduría General, Responsabilidad Fiscal, y Coparticipación a Municipios. Todos estos aspectos de temas gubernamentales se encuentran en formatos PDF y XLS, disponibilizados en distintos archivos para diversos períodos y categorías.

20. Santa Fe

La provincia de Santa Fe presenta en su sitio oficial de Datos Abiertos [GSF21], llamado “Datos Abiertos Santa Fe”, unos 139 conjuntos de datos (datasets), los cuales se encuentran enfocados en 10 categorías (“Administración Pública”, “Desarrollo Social”, “Economía”, “Educación y Cultura”, “Infraestructura y Obras”, “Justicia y Seguridad”, “Medio Ambiente”, “Producción”, “Salud” y “Territorio y Población”). 5 organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos fue desarrollado con Andino [AND21] y es gestionado con CKAN [CKA21], Versión: 2.6.3, y cuenta con filtros de búsqueda por nombre de dataset, Temas, Organizaciones y datos claves de etiquetas. Algunos de los tipos de formatos de archivos disponibilizados son: CSV, GeoJSON, XLS, SHP y PDF. Si bien el formato PDF no se considera abierto para interoperabilidad para software, también existen algunos archivos de este tipo.

21. Santiago del Estero

La provincia de Santiago del estero no presenta en su sitio oficial de gubernamental [GSE21], un Portal de Datos Abiertos. Cabe aclarar que no se encuentra información en su sitio sobre el contexto de Gobierno Abierto a los ciudadanos, pero si se disponibilizan algunos aspectos de Transparencia enfocados al ciudadano como información pública, como ser: “Licitaciones”, “Boletín Oficial”, y “Emisión del DNI”. Al igual que otras provincias, cabe aclarar que estos son datos públicos, no así datos abiertos.

22. Tierra del Fuego

La provincia de Tierra del Fuego presenta en su sitio oficial de Datos Abiertos [GTF21], llamado “Datos Abiertos | Antártida e Islas del Atlántico Sur”, unos 35 conjuntos de datos (datasets), los cuales se encuentran enfocados en 7 categorías (“Subsidio de gas”, “Socioeconómico”, “Obra Pública”, “Pensiones”, “Declaraciones Juradas”, “Ejecución Presupuestaria” y “Salarios”). Varias organizaciones son las que participan y colaboran con sus datos. Este portal de datos abiertos fue un desarrollado propio de la provincia, y cuenta con tipos de formatos de archivos disponibilizados en: CSV, ZIP, PDF y KML. Es un sitio sencillo y no posee filtros de búsquedas.

23. Tucumán

La provincia de Tucumán presenta en su sitio oficial de Datos Abiertos [GTU21], llamado “Conjunto de Datos”, unos 184 conjuntos de datos (datasets), los cuales se encuentran enfocados en 14 categorías (Identificadas cada una de éstas con un símbolo). 3 organizaciones son las que participan y colaboran con sus datos, siendo éstas: Poder Ejecutivo, Gobiernos Locales y Universidades. Este portal de datos abiertos fue desarrollado con Andino [AND21] y es gestionado con CKAN [CKA21], Versión: 298266510c80acf, y cuenta con filtros de búsqueda por nombre de dataset, datos claves de etiquetas, filtrado de organizaciones gubernamentales, ODS y Temas. Algunos de los tipos de formatos abiertos son: XML, CSV, XLS, XLSX, PDF, HTML, JPG (Otro), entre otros.

2.5. 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, como ser:

- a) CKAN [CKA21]: utilizado por Datos Abiertos Argentina;
- b) DKAN [DKA21]: utilizado por el Ministerio de Educación de Perú;
- c) Socrata [SOC21]: utilizado por ejemplo, por la Casa Blanca de EEUU.

Cada uno de 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 [PAN21]. El catálogo debe estar publicado en formato JSON o XLSX de estructura compatible, además, debe publicarse en una URL que cumpla alguna de estas 3 condiciones: a) Sea propiedad digital del organismo, b) 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 c) 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 documentan en su catálogo, todos los activos de datos digitales publicados en línea. Las pautas que deben cumplir según [PAN21] 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 que 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.

A continuación, se analizan los distintos principios internacionales relacionados con los aspectos a tener en cuenta para los datos abiertos.



2.6. Principios internacionales de los datos abiertos

Si bien existen muchos conjuntos de datos que son brindados en los portales de datos abiertos, tanto a nivel nacional como a nivel internacional, varias veces, éstos no poseen datos de calidad, es decir, no están en un formato correcto, existen errores ortográficos, o bien se encuentran campos incompletos entre otros aspectos. Para ellos, existen organizaciones que fomentan aspectos y/o principios de calidad en los datos.

La misión de Open Data Charter, Carta Internacional de Datos Abiertos [ODC21], es concientizar sobre los datos para que estén abiertos y disponibles gratuitamente para todas las personas y al mismo tiempo, esta organización tiene como misión proteger los derechos de las personas y las comunidades. Su visión general es que los gobiernos puedan recopilar, compartir y utilizar datos bien organizados y gestionados, para impulsar políticas y prácticas que permitan responder de manera eficaz, a los desafíos sociales económicos y ambientales más urgentes. Básicamente, trabaja en colaboración con más de 100 gobiernos y organizaciones que se esfuerzan para abrir datos basados en un conjunto de principios compartidos.

Otra de las entidades es Open Government Data, Datos de Gobierno Abierto [OGD21], que tiene como misión brindar principios similares para que los datos gubernamentales, y de esta forma se identifiquen como datos abiertos.

Si bien existen muchas entidades con fines similares, éstas son las más relevantes en aspectos internacionales, junto con [OGP12], las cuales apoyan la idea que los datos gubernamentales deben estar disponibles para todos los ciudadanos, debido a que son financiados con dinero público de éstos, por ende, deben estar a disposición en todo momento, como así también en un formato abierto, que respete estándares técnicos y permita su reutilización y difusión.

Para más información sobre los Principios brindados por estas entidades internacionales, consultar el [“Anexo B: Principios Internacionales de los datos abiertos”](#).

2.7. Aspectos de Calidad: Norma y Estándares

Un aspecto importante que ayuda a lograr la confiabilidad de los datos brindados o disponibilizados es la evaluación y certificación del nivel de calidad de los repositorios de datos organizacionales [GUA21].

Algunas de las normas que se enfocan en este tipo de aspectos de calidad son:

- ISO/IEC 25012 [ISO08], que define el estudio de las características de la calidad de datos entorno a la evaluación basada en los estándares internacionales.
- ISO/IEC 25024 [ISO24], que define medidas de calidad de datos para medir cuantitativamente la calidad de los datos en términos de características definidas en ISO/IEC 25012.
- ISO/IEC 25040 [ISO40], que define los requisitos y recomendaciones para la evaluación de la calidad del producto de software y aclara los conceptos generales. Además, ofrece una descripción del proceso para evaluar la calidad como así también, presenta el detalle de los requisitos para la aplicación de este.

Para analizar la calidad de datos, esta tesis se orienta en algunos aspectos de la Norma ISO/IEC 25012 [ISO08]. Esta Norma enfoca los puntos relevantes al modelo de calidad de los datos, y



especifica un modelo general para aquellos datos que se encuentran definidos en un formato estructurado dentro de un sistema informático. Puede utilizarse para detallar requisitos, establecer medidas, como así también para planificar y realizar evaluaciones de calidad de datos” [CAL19]. La calidad del producto de datos se puede entender como el grado en que los datos satisfacen los requisitos definidos por la organización a la que pertenece el producto. Son precisamente estos requisitos los que se encuentran reflejados en el modelo de Calidad de Datos mediante sus características (Exactitud, Completitud, Consistencia, Credibilidad, Actualidad, Accesibilidad, etc.). El modelo de Calidad definido por el estándar ISO/IEC 25012 se encuentra compuesto por las 15 características que se muestran en la siguiente Figura 2.4.

Existe una serie de medidas cuantitativas de calidad de datos que son universales con las cuales también se enfoca esta tesis [PWD21]:

- *Completitud*: es el grado en el que todos los atributos del dato están presentes.
- *Validez*: representa el ajuste de un valor de cada uno de los datos al conjunto al que pertenece.
- *Unicidad*: la medida en que todos los valores distintos de un elemento de datos aparecen sólo una vez.
- *Integridad*: tiene que ver con el grado de conformidad con las reglas de relación de datos definidas.
- *Precisión*: determina en qué medida los datos representan correctamente la verdad sobre un objeto del mundo real o se ajustan a lo establecido por una fuente autorizada.
- *Coherencia*: representa el grado en que una pieza única de datos contiene el mismo valor a través de múltiples conjuntos de datos.
- *Oportunidad*: este atributo de la calidad de datos permite conocer si éstos están disponibles cuando se requiere.
- *Representación*: tiene que ver con el formato, patrón, legibilidad y utilidad de los datos para su uso previsto.

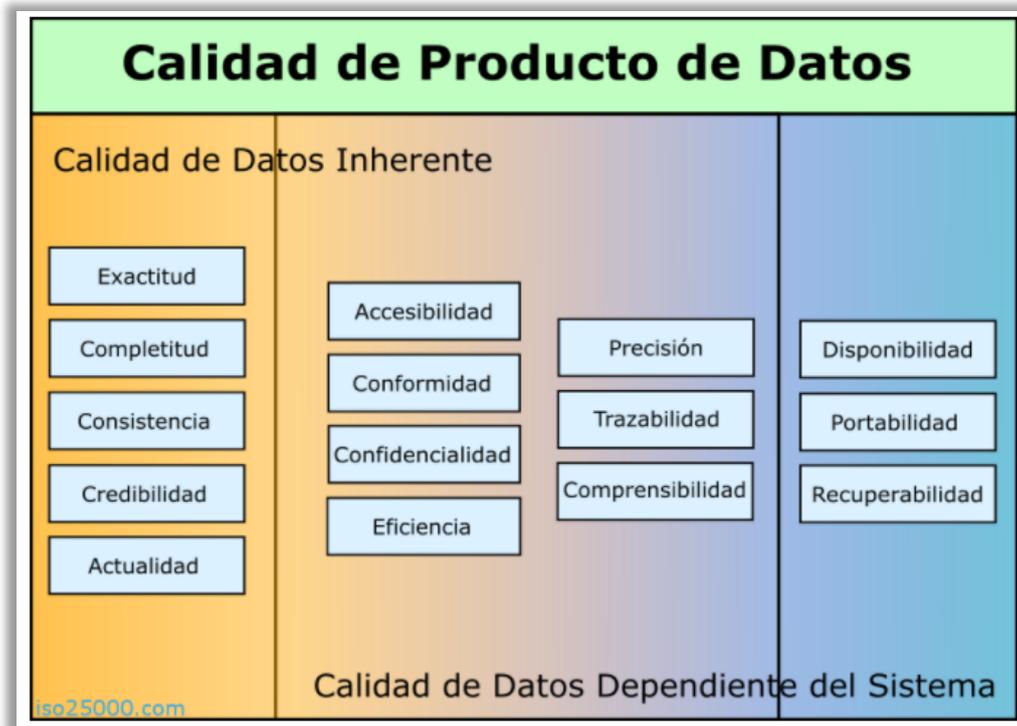


Figura 2.4 – Modelo de Calidad definido por el estándar ISO/IEC 25012. Fuente [ISO12].

Las características de calidad de datos están clasificadas en dos grandes categorías [ISO12]:

- **Calidad de Datos Inherente:** Se refiere al grado con el que las características de calidad de los datos tienen el potencial intrínseco para satisfacer las necesidades establecidas y necesarias cuando los datos son utilizados bajo condiciones específicas.
 - Valores de dominios de datos y posibles restricciones (Ejemplo: Reglas de Negocio gobernando la calidad requerida por las características en una aplicación dada).
 - Relaciones entre valores de datos (Ejemplo: Consistencia).
 - Metadatos.
- **Calidad de Datos Dependiente del Sistema:** Desde el punto de vista dependiente del sistema, la calidad de datos depende del dominio tecnológico en el que los datos se utilizan, y se alcanza mediante las capacidades de los componentes del sistema informático tales como: dispositivos hardware (Ejemplo: Respaldo software para alcanzar la recuperabilidad), y otro software (Ejemplo: Herramientas de migración para alcanzar la Portabilidad).

Para más información sobre el detalle del contenido de la Norma, se presenta el [“Anexo A: Norma ISO/IEC 25012”](#).

Como se trató en otras secciones, los portales de datos abiertos tienen diversos conjuntos de datos (datasets), los cuales son proporcionados a los ciudadanos con el fin de ser utilizados para relevamiento, análisis de datos, o bien acceder a éstos desde los distintos formatos abiertos (CSV,



JSON, XML, etc.) para permitir una mejor interoperabilidad entre software y/o persistencia en bases de datos. Esto logra una manera más apropiada de brindar información y conocimiento en el marco gubernamental, apoyando así, el paradigma de transparencia en la Administración Pública Nacional (APN).

Para cuestiones referidas a calidad de los datos, siempre se deben definir métricas que son los instrumentos utilizados para medir un determinado aspecto o factor, como, por ejemplo: medir de manera cuantitativa la cantidad de registros duplicados en un dataset. En el siguiente capítulo se muestran los trabajos relacionales que se presentan en la actualidad para este contexto.



Capítulo 3 – Trabajos Relacionados

En esta sección, se lleva a cabo una visión más sistemática de los temas investigados en los contextos triviales que involucran a esta tesis, es decir, se analizan los distintos trabajos relacionados que se organizan en diversas áreas de interés, en base a estudios elaborados en los últimos años, como ser: desde la perspectiva del desarrollo y diseño de los repositorios de datos abiertos gubernamentales, cuestiones técnicas en lo relacionado al desarrollo y despliegue de portales de datos abiertos, como así también, cuestiones concernientes en la usabilidad de los sitios web de los portales abiertos.

Como otra sección, se estudia la calidad en la construcción de los portales, por lo que se enfoca en algunas investigaciones relacionadas a métricas y modelos propuestos para este contenido.

Además, se exponen diversos trabajos relacionados a la evaluación de la reutilización de estos datos abiertos públicos y la influencia que tienen en los ciudadanos. Finalmente, se presentan trabajos orientados al enfoque del crowdsourcing y cómo puede beneficiar esta temática en la gestión de portales de gobierno abierto para los ciudadanos involucrados.

Cabe destacar que, para el proceso de búsqueda de esa sección, se analizaron diversas fuentes administradas desde la utilización de la plataforma *Mendeley*, siendo éstas:

- Repositorio digital *IEEE Xplore*;
- Repositorio digital *SpringerLink*;
- Repositorio digital *Scopus*;
- Repositorio digital *Science Direct*.
- Repositorio digital *ACM*.
- Repositorio digital *Google Scholar*.

A continuación, se presentan las secciones por temas relevados de los trabajos relacionados indicados anteriormente.



3.1. Construcción de Bancos de datos abiertos

Para esta sección se tomó en cuenta la investigación de trabajos relacionados a la construcción del armado de bancos/repositorios de datos abiertos, con el fin de utilizarlos para el empoderamiento ciudadano.

Algunos de los autores hacen referencia a la incorporación de los beneficios del nuevo paradigma de datos abiertos, en base a propuestas de modelos, como ser Rojas [RRL17], en el que integró algunas de las soluciones de inteligencia de negocios, para la construcción de los portales abiertos, y su propuesta se destacó en administrar correctamente los portales de datos abiertos que se encuentran a disposición de los ciudadanos, a través de la utilización de un DSL (Lenguajes de Dominio Específico) que utiliza un enfoque temático y permite la definición de una sintaxis abstracta y concreta de datos para el contexto de Gobierno Abierto. Otros autores [BYF16], proponen realizar cruces con estos datos abiertos públicos disponibles, lo que permite un análisis de campo con varios recursos, como, por ejemplo, un estudio realizado para detectar villas y asentamientos en el Partido de La Matanza, provincia de Buenos Aires, Argentina, mediante teledetección y un sistema de datos abiertos geolocalizados en base al relevamiento realizado por una organización no gubernamental en el año 2013. La metodología de este trabajo tuvo que ver en una etapa de obtención y preprocesamiento de datos e imágenes, segmentación de las imágenes, cálculo de atributos y aplicación de un algoritmo de clasificación de segmentos previamente definido. Por lo que, la propuesta es una metodología que reduce el área del territorio a relevar, teniendo en cuenta la periodicidad y actualización de los conjuntos de datos abiertos (datos censales, viales, naturales georreferenciados, imágenes satelitales y algoritmos de aprendizaje automático). Esa capacidad de reutilización y accesibilidad de datos abiertos, aborda distintos estudios, como el de Venegas Álvarez [VEG18], en el que profundiza sobre los posibles tipos de usos que los usuarios pueden dar a los datos abiertos, por lo que sostiene que se puede diseñar un portal de datos abiertos, orientando a una metodología más genérica enfocada al nivel de apertura, capacidad de reutilización y accesibilidad de los datos de gobierno, particularmente de transparencia, considerando en su estudio, casos como el de los países de Chile, Costa Rica y Uruguay. Dichos autores también mencionan la necesidad de definir vocabularios y publicar datos enlazados puede no resultar natural en un principio, por lo que es importante, identificar quienes tendrán que realizar dicha tarea al diseñar un portal abierto, ya que este vocabulario es relevante para facilitar la integración de fuentes de datos distintas, como son los datos publicados por la transparencia activa. La apertura de los datos públicos crea una oportunidad para todos los interesados, incluidos los individuos, las empresas privadas y las organizaciones públicas, para generar un valor sostenible para la sociedad en general, como se desarrolla en el trabajo realizado por Thorhildur Jetzek, Michel Avital y Niels Bjorn-Andersen [JET19], en el que presentan una propuesta de mejora en los procesos administrativos gubernamentales para que sean más eficientes. Su estudio, realiza una prueba y validación de un modelo con datos cuantitativos secundarios de 76 países tomando como base el Open Data Barometer [ODB21], publicado por el Open Data Institute [DSF21], analizando la recolección de datos y operatoria de variables de investigación, como ser variables para: la apertura de datos, gobernanza digital, cuantificación del liderazgo digital del gobierno, marcos regulatorios de protección de datos y variables de privacidad, como así también, variables relacionadas a la infraestructura digital colaborativa.

En cuestiones de identificación de las distintas dificultades que pueden surgir en la construcción de los repositorios de portales de datos abiertos, existen trabajos de estudio, como ser el de Jędrzej Wieczorkowski [WIE19], en el que menciona las barreras clásicas que se deben analizar



para el desarrollo y administración de estas, siendo: barreras técnicas, económicas, organizativas y humanas. Pero además de estas barreras históricas, en el mundo de datos abiertos, se debe tener en cuenta los requisitos de infraestructura para el procesamiento de grandes volúmenes de datos, por lo que es necesario realizar un foco especial en los conceptos de eficiencia de procesamiento, del análisis de flujos en tiempo real de los datos y de un correcto desarrollo de herramientas de visualización. En cuestiones técnicas, lograr interpretar la correlación entre grandes volúmenes de datos heterogéneos requiere el conocimiento adecuado, por lo que los científicos de datos, ya que tienen un papel fundamental en este rol. Además, todo lo anteriormente comentado, debe llevarse a cabo con un cambio de metodología por parte de los gerentes y adaptarla a la nueva cultura organizacional: la cultura de los datos, que se relaciona con la barrera organizativa. Esto da lugar, a pensar en dos nuevas barreras más que menciona el autor, siendo la barrera legal y la del dato puntualmente. Para la legal, se debe tener autorización del organismo de origen de los datos, con el fin de evitar la publicación y el acceso a los datos que no son autorizados y pueden ser de carácter confidencial o bien sensibles para un organismo público, básicamente, pueden existir datos que pertenecen a personas y organizaciones de inteligencia, los cuales pueden estar sujetos a otras restricciones. Este concepto de big data para el mundo de datos abiertos, requiere de una escala adecuada de procesamiento de datos, conexión de datos y acceso a datos históricos definidos. Además, es vital pensar en el análisis de verificación de la calidad de este tipo de volúmenes de datos, ya que la cantidad de datos puede afectar la calidad de éstos, es decir, este concepto de big data contradice en parte, al principio universalmente aceptado, que indica que la baja calidad de los datos de entrada puede dar como resultado de salida, una mala calidad (regla GIGO: basura adentro, basura afuera) [WIE19]. Sin embargo, dependiendo de la aplicación, existe un nivel mínimo aceptable de calidad de datos utilizados. Es por ello por lo que, se pueden tomar dos enfoques básicos: cuidar la selección de datos de alta calidad, es decir, limitar la cantidad de datos fuentes o bien ignorar parcialmente el problema de la calidad de los datos fuentes. En resumen, se puede aumentar la complejidad de los procesos de recopilación de datos. Sin embargo, en la práctica, se encuentran soluciones de big data que tienen requisitos mínimos de calidad. Otro de los puntos que hace mención esta investigación, es qué en Polonia, los informes y programas gubernamentales indican que los usos de formatos de datos públicos son de baja madurez, es decir, no estructurados, con licencia, y que no admiten datos vinculados, y el uso insuficiente de APIs no colabora en un mejor desarrollo de los portales abiertos. Esto puede estar relacionado con la imperfección de la ley, siendo la ley de reutilización la que no obliga al uso de métodos maduros y útiles de publicación de datos abiertos.

En base a lo relevado, es muy importante que el estado de cualquier país brinde datos útiles de calidad adecuada y, además, defina principios legales que permitan respaldar el uso práctico de los datos abiertos, como así, el desarrollo de APIs, para un apropiado proceso de adquisición de datos y acceso de éstos, en formatos de diferentes fuentes vinculados.

Por otro lado, se releva que si bien existen posibles problemas en los repositorios centrales de información pública en varios países del mundo [WIE19], es importante identificar conjuntos de datos disponibles con estándares de datos abiertos, como ser el estándar “rdf”, el cual permite el uso de datos vinculados y utilización de APIs. Además, es necesario analizar los métodos de intercambio de datos, para un correcto procesamiento y uso de datos abiertos vinculados. El trabajo de Ning Zou Robert Gradeck [XIA18], describe un estudio para investigar las misiones, servicios y tareas operativas proporcionadas por cuatro centros de datos de gobierno abierto: NYC OpenData (Nuevo York Open Data Center), DataSF (portal de datos abiertos de San Francisco), WPRDC (Centro de Datos Regional del Oeste de Pensilvania) y el almacén de datos de Londres (portal de datos abiertos del Gran Londres). En todos los casos, la función vital consistía



en mejorar el uso de los datos, y en estos estudios, se abocó a brindar datos sobre servicios y tareas en cinco categorías comunes: *Disponibilidad* (garantizar que los datos estén abiertos, los datos del gobierno deben estar disponibles para cualquier persona que quiera acceder a ellos y utilizarlos, es decir, fácilmente detectables y descargables), *Comprensibilidad* (proporcionar más información (como información contextual) o herramientas, para ayudar a los usuarios a comprender los datos y utilizarlos), *Ayuda Técnica* (se centra en la creación de API o herramientas para servir mejor a varios tipos de usuarios de datos para usar y aplicar datos), *Compromiso Social* (consultas o comentarios de los usuarios también pueden ayudar al gobierno local a mejorar la prestación de servicios) y *Mejorar la Alfabetización de Datos de Usuario* (potenciar el uso de datos, por ejemplo: capacitación y respuesta a preguntas de usuarios).

Para estos portales abiertos, fue importante tener algunas funcionalidades básicas, como ser: a) Intérprete de datos, b) Consultor de datos y c) Asistente visual de datos. Estas 3 características, se podrían ampliar y aplicarse en otros portales de datos abiertos y centros de datos basados en dominios, por ejemplo, repositorios de datos de investigación, para optimizar la entrega de datos abiertos, y así, facilitar un mayor valor del intercambio de datos. Como punto final de esta investigación, el autor sugiere que es importante acompañar a la administración de estos repositorios, por lo que propone funciones o puestos especializados de datos de gobierno abierto: a) Intérprete de datos: este rol es para trabajar con proveedores de datos para crear guías de datos, recopilar y crear blogs e historias de datos relacionados con datos, trabajar de forma independiente o con programadores para crear mapas u otras visualizaciones, e informar políticas relacionadas con los datos; b) Consultor de Datos: ayudar directamente a los usuarios a comprender los datos y enseñarles habilidades técnicas para utilizar los datos con precisión; y c) Asistente visual de datos: para ayudar a los usuarios a visualizar o manipular datos mediante el desarrollo de herramientas que pueden ser utilizadas directa y fácilmente por éstos. Es decir, los usuarios pueden utilizar las herramientas gráficas y las aplicaciones para crear las visualizaciones específicas que necesiten.

Es importante hacer foco en los problemas que hoy existen en la mala interpretación de los datos, ya que muchas veces, no alcanza con un diccionario de datos de los datos abiertos, esto se debe a su complejidad para algunos casos, y sobre todo si los datos son de organismos en los que es necesaria una lógica más interpretada del tema. Es por ello por lo que, algunos estudios sugieren arquitecturas de software para estructurar la información en datos abiertos de repositorios distribuidos ubicados en los sitios web de los gobiernos locales, como la investigación de Andreiwid Sheffer Correa [COR18], para evitar algunos problemas de interpretación de datos abiertos. Por lo que este trabajo sigue la metodología de Diseño Dirigido por Atributos (ADD), donde el análisis de requisitos para la arquitectura propuesta se realizó en base a problemas de disponibilidad ocurridos en 561 sitios web de ciudades brasileñas. El propósito de la arquitectura es proporcionar artefactos arquitectónicos para implementar un sistema colaborativo para estructurar la información en datos abiertos para permitir el uso y la reutilización de los datos por cualquiera y para cualquier propósito. La arquitectura se desarrolló mediante la identificación de problemas con respecto a la disponibilidad de datos que se encuentran en los sitios web del gobierno local en los municipios brasileños, el enfoque de colaboración propuesto permitió desarrollar mecanismos de evaluación y clasificación para formar una red de contribuyentes autosostenida y moderada. Otro trabajo, como el de [OEF18], analiza la importancia de que los repositorios de datos abiertos deben actualizarse periódicamente y deben brindar calidad en sus datos expuestos, por lo que sugiere agregar como parte de un procedimiento gubernamental, posibilidades de brindar solicitudes de pedidos de los ciudadanos a las distintas agencias



gubernamentales que determinarán, si los datos pueden o no ser liberados, y en caso de considerarse como datos públicos.

Otro problema comúnmente asociado con los datos de gobierno abierto es el tema de los formatos de datos propietarios, lo que hace que el uso sea muy laborioso y propenso a errores. Por lo que se hace importante, la estandarización entre las diferentes fuentes de datos, lo que es muy difícil conciliar. Este es un tema mucho más grande y difícil de abordar, porque los datos se recopilan de varios sistemas de información diferentes, que son propiedad de diferentes áreas del gobierno, y en cada uno de ellos, existen normativas distintas de sus propias secretarías, departamentos, agencias, etc. Esta accesibilidad y conciliación de diversas fuentes, permitirá que los portales de datos abiertos, que sean más reconocibles, accesibles, y disponibles en formatos alternativos y con esquemas de licencias que permitan la reutilización gratuita, como sugiere el trabajo de realizado por EE. UU. [PYO18], el cual indica que diversas partes interesadas, podrán desarrollar aplicaciones de datos innovadoras y así, obtener un mejor análisis de estos repositorios bien desarrollados y definidos. Esto, además, mejorará las habilidades técnicas y la capacitación de los usuarios/ciudadanos para su mejor utilización. Para avanzar en la adaptación de este nuevo paradigma, el autor menciona métodos para aumentar el conocimiento de gobierno de datos abiertos, y menciona como España, Italia y Estados Unidos han tratado de mejorar las habilidades de los usuarios y los han comprometido en el uso de este contexto. Para ello, algunos de los hallazgos analizados en estos países son las habilidades de introducción y análisis, que se les enseñan a los ciudadanos, para fomentar el uso de datos abiertos, es decir, se presentan sus beneficios en su uso, y se completan sus conocimientos en la incorporación de las intervenciones de formación en los contextos específicos de intereses y expectativas puntuales que les interesan a los diferentes tipos de usuarios, lo que lo hace fundamental para el éxito y el compromiso de éstos.

Algunos estudios [PYO18] apuntan a mejorar las habilidades de los usuarios de estos portales, dependiendo los contextos alternativos de datos gubernamentales que se visualicen, por lo que algunas investigaciones proponen tipos de usuarios de los portales de datos abiertos, ya que cada tipo de usuario tiene diferentes intereses y uso previsto. Por ejemplo, los empleados gubernamentales pueden utilizar los datos para mejorar los servicios públicos, así como los procesos de toma de decisiones y políticas. Otros pueden ser usuarios innovadores, que incluyen programadores y desarrolladores o bien investigadores, periodistas de datos y/o activistas, que intervienen con el fin de crear conocimiento. Otro de los puntos que marca este trabajo, es que los ciudadanos casi nunca consumen los datos abiertos directamente, sino que utilizan los datos con mayor frecuencia a través de uno o más mediadores. Para ello, el Open Data Institute (ODI) [DSF21], desarrollo un framework, que incluye varias habilidades básicas, como ser: introducción a los datos abiertos, conceptos de publicación mediante plataformas para mejorar la calidad, elegir formatos, vincular datos, limpiar datos e impulsar la usabilidad, por otra parte, impulsar a comunidades para innovar con los datos abiertos. Es decir, esto permite mejorar las habilidades y los conocimientos de los usuarios mediante iniciativas de formación y participación.

Actualmente, con el fin de que los datos sean comprensibles para los ciudadanos, las plataformas han comenzado a proporcionar informes de análisis de datos, para ayudar a los usuarios a obtener información sobre los datos. Algunos centros de datos han generado guías de usuario además de proporcionar metadatos sobre sus conjuntos de datos y un diccionario de datos, para mejorar la educación en el manejo de los datos abiertos, de los usuarios/ciudadanos, algunos de ellos cuentan con soporte técnico para responden preguntas o consultas. Es decir, todo este contexto, ayuda a mejorar el uso de los datos abiertos públicos.



3.2. Evaluación de la reutilización de datos abiertos

En esta sección se analizan los trabajos relacionados a la reutilización de los datos abiertos, es decir, se identifica el grado de uso de los datos abiertos públicos por parte de los ciudadanos, como de los organismos privados y/o públicos.

Las implementaciones de los datos abiertos en la política pública pueden ser analizados desde el punto de vista de la usabilidad en los portales de datos abiertos, básicamente son la base del modelo del gobierno abierto, por lo que existen estudios [RJA19], que presentan enfoques que permiten la intervención de agentes intermediarios que facilitan el amplio aprovechamiento de los datos abiertos de gobierno. En base a este contexto, se pueden tomar variables que inciden sobre la decisión de involucrar intermediarios o no en estos contextos de gobierno abierto. Por lo que, existen mediciones. Es decir, se identificó que, en la ausencia de un auténtico compromiso de apertura del gobierno y una sólida participación ciudadana, el modelo del gobierno abierto puede quedar en una promesa sin resultados concretos. Para ello, algunas de las propuestas, es incorporar a un intermediario basado en servicios de auditoría y de análisis de información, que pueda ayudar en el proceso de inculcar la cultura de gobierno abierto a los ciudadanos. Este análisis de acceso a la información [VPJ17] se puede enfocar desde el lado de la reutilización de los datos, y así utilizar métricas para evaluar si éstos son o no reutilizados desde el acceso de portales de datos abiertos. Existen algunos trabajos [AAD19] que proponen prototipos de software para la evaluación de principios de datos abiertos y que permiten, además, validar el cumplimiento de leyes de los datos abiertos, como estudios recientes que se realizaron en los portales abiertos de Colombia, a partir de los principios establecidos por el Open Government Data [OEC20]. Por lo que, analizar los principios de los datos abiertos y las leyes que involucran políticas sobre el acceso a la información, es importante, para la implementación de una arquitectura que pueda evaluar los datos recolectados de plataformas de datos abiertos para establecer la madurez de este tipo de información. Otras propuestas [MSS20] se enfocaron en la evaluación de portales de datos abiertos, en base a criterios definidos por los autores, como ser el planteamiento de metodologías e indicadores que miden la calidad de los portales y de sus datos, para los portales de datos abiertos en los países de España, Brasil, Costa Rica, Taiwán y la Unión Europea. De este estudio se presentan dos enfoques bien marcados:

- a) Datos publicados: abarcando la calidad (identifican factores de disponibilidad, actualización, accesibilidad, visualización, formatos de publicación y completitud), uso (identifican los 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 (identifican el factor de búsqueda, Navegabilidad y uso/consumo/descarga de datos) y mecanismos de comunicación (identifican los factores de comentarios y discusión, fuente-usuarios y solicitud).

Cada metodología aporta un enfoque distinto en la medida en que están planteados sus criterios de evaluación, lo cual puede llevar a que el elemento estudiado (portal o datos) tenga niveles de calidad distintos, según la metodología que se haya empleado.



Por otro lado, algunos trabajos [ZNZ19] realizan foco en nuevos modelos de medición del nivel de uso de datos abiertos del gobierno abierto, con 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 [ODB21], con el fin de medir el éxito y el impacto de las iniciativas de datos abiertos de los gobiernos en todo el mundo. Al medir la intención conductual de usar tecnologías, los investigadores adaptan teorías y modelos relacionados con ese objetivo. En cuanto a la evaluación de la usabilidad [MAC18], se realizaron propuestas de buenas prácticas, en base a la investigación de la utilización de datos para los países de Australia, Canadá, India, Estados Unidos y Reino Unido, para mejorar la capacidad de las partes interesadas para descubrir, acceder y reutilizar estas fuentes de información en línea.

Los portales de datos abiertos suelen proporcionar información sobre un conjunto de datos en forma de una descripción con metadatos y permiten el acceso directo a los conjuntos de datos a través de la descarga y, además, a través de la interfaz de programación de aplicaciones (API). Esto ayuda a los usuarios a buscar conjuntos de datos relevantes o a examinar conjuntos de datos por categorías, etiquetas, organizaciones o formatos. Existen sistemas de gestión de datos libres y abiertos, como la Red Integral de Archivos de Conocimientos (CKAN) [CKA21] o Drupal Knowledge Archive Network (DKAN) [DKA21] que han sido adoptados por muchos gobiernos. Algunos portales proporcionan características adicionales, API o puntos de conexión SPARQL, listas de aplicaciones creadas con un determinado conjunto de datos, características de discusión y comentarios y herramientas integradas para visualizar datos, como se indicó en trabajos anteriores. Los autores proponen una lista de criterios para la evaluación de la usabilidad de los portales de datos abiertos: a) Dimensión de especificaciones del conjunto de datos (descripción, editor, categorías y temáticas, fecha de lanzamiento y actualización, formatos legibles, licencia de datos abiertos visualización y herramientas de análisis); b) Dimensión de comentarios del conjunto de datos (Documentación, foros, calificación de usuarios, comentarios y redes sociales), c) Dimensión de solicitud de conjunto de datos abierto (Formulario de solicitud, lista de solicitudes y participación del proceso).

Otro de los enfoques que se encuentran establecidos en la actualidad, es el estándar global para el gobierno de los recursos de petróleo, gas y minerales. Esta es una iniciativa de transparencia de las industrias a nivel internacional (EITI, Extractive Industries Transparency Initiative) [EIT21] para fomentar el análisis de los datos publicados (datos abiertos) por los países implementadores del modelo EITI. La utilización de los principios que ofrece EITI facilita una mejor comprensión pública de los ingresos y gastos del gobierno a lo largo del tiempo, lo que permite ayudar al debate público e informar la elección de opciones adecuadas y realistas para el desarrollo sostenible.

3.3. Crowdsourcing y datos abiertos

En esta sección se analizan algunos de los trabajos relacionados con el concepto de la metodología de crowdsourcing y los datos abiertos.

Algunos de los autores [CAC17], [GRA16] proponen en su trabajo de investigación un proceso de construcción de mecanismos de datos y colaboración abiertos distribuida (crowdsourcing), como elemento esencial en la participación y cultura ciudadana en la ciudad. Esto se debe a la posibilidad de fomentar el uso didáctico de las tecnologías de la información y las comunicaciones TIC, por lo que, en dichas propuestas, se encuentran diseños prototipos de software que permiten



a los usuarios colaborar activamente en la construcción del conocimiento del gobierno abierto. Por ejemplo, a través de aporte de ideas e información sobre problemas relacionados con alumbrado público, movilidad, espacio público, infraestructura y limpieza pública, los cuales son expuestos públicamente a través de una plataforma abierta con acceso a través de formatos XML, RDF y OWL.

Para este tipo de portales de datos abiertos colaborativos, se debe seguir un proceso específico, con el fin de garantizar la correcta difusión de los datos a través de plataformas que debe ser supervisadas y gestionadas por personal gubernamental, por lo que es vital que ésta sea administrada por personas idóneas.

En base al crowdsourcing ciudadano, se pueden encontrar ejemplos de plataformas de este concepto, como ser:

- Challenge.gov [CHA21]: es un ejemplo de una experiencia con un recorrido notable dentro de la Administración Pública en los Estados Unidos. Es una propuesta directamente del expresidente Obama, sobre crear una plataforma de apoyo a la estrategia de gobierno abierto mediante un instrumento de política pública denominado Concursos y Premios. Su fin es promover competencias para la identificación de soluciones a problemas sociales e incorporar el conocimiento existente fuera de las organizaciones públicas.
- SantanderCityBrain[SAN21]: se trata de una plataforma social desarrollada por el Ayuntamiento de Santander, dirigida a todas las personas, sean o no vecinos de la ciudad, que deseen compartir propuestas, comentar y votar las ideas y proyectos que compartan otros usuarios, con el fin de colaborar en el desarrollo de la ciudad, en general, y la ejecución de su plan estratégico.

Por otro lado, algunos ejemplos de plataformas sociales de colaboración abierta entre empleados públicos pueden ser:

- GitHub [GIT21]: se ha convertido en una plataforma de creación colaborativa de software en la que participan activamente cada vez más administraciones públicas y profesionales relacionados con el sector público. Esta comunidad social de programadores se centra en la reutilización del código creado por otras organizaciones previamente, haciendo cambios para su adaptación al contexto propio de aplicación. Básicamente, el autor comenta que la colaboración no solo se produce para crear partes de código fuente (elementos de páginas web, o plataformas de datos abiertos) sino que, además, los equipos de investigación gubernamentales también comparten sus datasets, y algoritmos, y recientemente agencias seleccionadas han comenzado a desarrollar documentos de política pública en formato texto dentro de GitHub. La plataforma ofrece opciones de interacción social, de manera que las organizaciones y usuarios pueden crear un perfil propio, en el que introducir una fotografía, nombre completo, localización, dirección de correo electrónico y organización a la que pertenecen o a las que se vinculan dentro de la plataforma.
- NovaGob [NOG21]: Es una plataforma social, especializada en el sector público en la que interaccionan sus profesionales de una manera abierta y sin ningún tipo de restricción organizativa. Permite conectar a los empleados públicos de distintos países e incluye a profesionales de todos los niveles de gobierno con el propósito de que desarrollen



proyectos de manera abierta y sin el control externo de sus organizaciones de pertenencia.

En todos ellos, se ha demostrado la existencia de nuevos principios, normas, herramientas y prácticas para el funcionamiento del sector público, pasando de modelos jerárquicos, basados en reglas formales hacia un esquema de funcionamiento horizontal, con reglas más flexibles y comunidades interorganizativas. Las iniciativas del gobierno abierto proporcionan a los interesados, medios para obtener información gubernamental sobre una localidad o país, para reutilizarlos y crear una fuente de enriquecimiento de varias maneras: nuevos servicios para usuarios, desarrollo económico y mayor transparencia. Algunos estudios, [SIN19] identificaron que, si bien los portales gubernamentales de datos abiertos ofrecen herramientas para presentar, buscar, descargar y visualizar la información gubernamental, existen algunos problemas de disponibilidad, o bien del tipo de redundancia de datos en los portales abiertos. Por lo que, se proponen frameworks para generar datos gubernamentales inteligentes vinculados abiertos, este marco se beneficia de varias tecnologías, datos vinculados, procesamiento de lenguaje natural para agregar en un nivel semántico conjuntos de datos similares y sistemas de recomendación basados en calificaciones para proporcionar sugerencias de conjuntos de datos que pueden representar un interés potencial para los ciudadanos. Este trabajo parte del concepto de la visualización de red de un conjunto de datos: un diagrama de red colaborativa, donde los conjuntos de datos están representados por nodos.

Algunos estudios [SAN17], [SON18], demuestran que, en comparación con la externalización de software tradicional, las aplicaciones actuales basadas en datos abiertos, crowdsourcing y Web 2.0 requieren interdependencias más complejas. Es decir, cuanto más crítico sea un servicio gubernamental, más interdependencias colaborativas deberán existir. Por otra parte, se pueden proponer evaluaciones simples, esto es, a través de la colaboración de los ciudadanos, se analizan portales de datos que permiten a los usuarios recuperar sus conjuntos de datos con opciones de búsqueda, incluidas palabras clave, tipos de datos o etiquetas generadas por el usuario, etc. Sin embargo, se limita a descubrir conjuntos de datos basados en su contenido. Por esta razón, los usuarios deben verificar si estos conjuntos de datos son adecuados para sus propósitos sobre los resultados de búsqueda. Para resolver este problema, se recurre a la búsqueda semántica, analizada en los trabajos relacionados de la sección anterior.

Otro de los trabajos realizados [GIL17], exploran el papel que desempeña el personal de datos abiertos del gobierno municipal en el movimiento de datos abiertos, y cómo sus funciones emergentes pueden aumentar los beneficios sociales y económicos de los datos abiertos, así como aumentar la participación ciudadana. Por lo que se analiza la forma en que las iniciativas gubernamentales de datos abiertos dan forma a la interacción entre los gobiernos y los ciudadanos. Estos incluyen no saber quiénes son los usuarios de datos abiertos, la naturaleza jurisdiccional de la forma en que se administran los datos y la forma en que los gobiernos perciben lo que constituye datos abiertos.

Si bien el crowdsourcing es un método cada vez más común de prácticas de gobierno abierto para fortalecer la democracia participativa, su impacto en la gobernanza no está claro. Es decir, aunque el crowdsourcing pretende crear igualdad e inclusión en la formulación de políticas, hallazgos revelan que la sobrecarga de datos cívicos y la jerarquía de filtros complican la adopción del crowdsourcing como una innovación democrática en la gobernanza. Al aplicar el crowdsourcing digital a la formulación de políticas, los gobiernos locales y nacionales buscan recopilar datos, mejorar las políticas e involucrar en un mayor grado a los ciudadanos. Es por esto por lo que, se



pueden analizar los comentarios de los ciudadanos y calcular las palabras más utilizadas, para sacar una estimación de los temas más relevantes, esto lleva a estudios que demuestran [BAR19] evidencia empírica para responder al problema revelado de cómo medir el grado de respuesta del gobierno a las sugerencias de la multitud y cómo identificar los factores que determinan dicha capacidad de respuesta.

En lo que respecta a las arquitecturas propuestas para esta temática [FRR17], [RUI20], [HLA19], [CIT20], se orientan a los servicios, y está compuesta por varios sistemas con propósitos específicos. Algunos de éstas, se orientan a los repositorios de documentos, donde se almacenan todos los datos de fuentes gubernamentales, y esta base de datos, se alimenta a través de procesos ETL que sirven a sistemas que componen la plataforma de portales abiertos.

3.4. Calidad en los portales de datos abiertos

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

Uno de los aspectos fundamentales mencionados por la Comisión Económica para América Latina y el Caribe, CEPAL [CEP21], una de las cinco comisiones regionales de las Naciones Unidas, es que los gobiernos recaudan una gran cantidad de datos de alta calidad. Normalmente, esto se traduce en que el Estado se convierta en un poderoso monopolio de datos capaz de estructurar y homogeneizar las interacciones entre él y sus ciudadanos. Estas interacciones unilaterales son costosas y no responden necesariamente a las necesidades de los ciudadanos, lo que limita innecesariamente las actividades gubernamentales [BIB21]. Por ello, es necesario contar con técnicas y estrategias para verificar la calidad de los datos que se brindan a la comunidad, y de esta manera obtener un análisis que ofrezca un valor agregado a la sociedad. Uno de los puntos que aborda el informe del Banco Interamericano de Desarrollo, BID, sobre datos abiertos en América Latina y el Caribe [MUE18], es que la apertura de los datos genera una mayor interacción entre gobiernos y ciudadanos. y, además, posibilita una instancia de cocreación con otros actores que conforman la demanda de datos como investigadores, emprendedores, sociedad civil, desarrolladores, periodistas y otras instituciones gubernamentales. Además, explica que las agencias deben 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.

Existen trabajos [MMO16], [BEM17], [IGL19], [RRL17] [ARI16] que se enfocan en establecer criterios y clasificaciones de niveles de calidad de los datos abiertos. Algunos escenarios de análisis se orientan a partir de la reutilización de datos abiertos y públicos. Por otra parte, algunas técnicas de medición de calidad se basan en el concepto de la disponibilidad de los datos en los portales de datos abiertos, con el fin de propiciar un nivel de calidad adecuado para los consumidores de estos. Además, se 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. Otros trabajos [AOU18], orientan la evaluación de la calidad de los portales abiertos en el análisis del modelo de cinco estrellas de Berners-Lee [OPD12] y otros factores que ayudan a evaluar la calidad de los datos para la reutilización. Por otro lado, proponen que se puede analizar



la relación entre la demanda de los datos abiertos y la dispersión de distribuciones y conjuntos de datos disponibles sobre una determinada temática, por lo que, es interesante estudiar el resultado de si concentrar la información en pocos conjuntos de datos más completos y manejables, podría ayudar a mejorar la eficiencia en la publicación de los datos.

Otros trabajos [NAY18] proponen técnicas para resolver los problemas con los sitios de gobierno de datos y para ello, desarrollan framework o técnicas que proporcionan servicios de datos basados en la web que se estructurarán semánticamente. Este framework puede contemplar aspectos de: a) Proveedor de datos: Que los proveedores de datos alojen sus datos de manera estándar para que, en el futuro, otros puedan usar / reutilizar los datos de manera más semántica, donde el proveedor de datos aloja los formatos de conjuntos de datos estructurados como CSV, XML, JSON, etc.; luego, que los datos estén disponibles utilizando un esquema de metadatos o un vocabulario para metadatos. Para mejorar la calidad del software de los portales que realizan un tratamiento de datos abiertos geoespaciales, existen trabajos [LAF18], que proponen enfoques que es el desarrollo y aplicación de un vocabulario base transversal que apoya la descripción de temas geoespaciales. El método de anotación resultante está integrado en una nueva plataforma de colaboración de acceso abierto admite la difusión pública de datos públicos. Para ello, hacen mención de la utilización de ArcGIS Hub [ARC21] que es una nueva plataforma de acceso abierto que admite y organiza el compromiso cívico y la colaboración directa entre los gobiernos y sus ciudadanos. ArcGIS Hub posee nuevas funcionalidades para compartir datos abiertos, catálogos de metadatos configurables, integración con registros de metadatos regionales y nacionales.

Por otra parte, trabajos desarrollados para una tesis doctoral [CAD19], enfoca una propuesta de marco de referencia para la publicación de datos abiertos comprensibles basados en estándares de calidad, el cual está compuesto por tres niveles de desarrollo, como ser: estratégico, táctico y operativo. Cada uno de estos niveles estudia una serie de procesos que deben realizarse para completar un procedimiento de publicación de datos abiertos teniendo en cuenta criterios de comprensibilidad. Además, investigaciones diversas [FRA16] plantean métodos centrados en el usuario para medir la calidad de datos abiertos. Puntualmente en este caso, se proponen métricas para evaluar la calidad de datos en base a la necesidad de los usuarios. Otros modelos de investigación innovadores [FIT19], [KUB18], destacan que la calidad tecnológica y los factores de confianza de la veracidad de los datos abiertos en los portales de gobierno abierto, depende específicamente de la actitud del usuario, es decir, se apunta a la intención de utilizar datos abiertos en los portales web, ya que de esta interacción de los usuarios/ciudadanos, se pueden analizar factores como: la confianza de la información brindada en un sitio web de datos, la calidad de la información en lo que respecta a la utilidad que puede tener, y la calidad del sistema en cuanto a la usabilidad.

Varios Gobiernos y organizaciones que fomentan el Gobierno Abierto [FDT21] [ODS14] [DGE17] [OBV15], publicaron guías de buenas prácticas y/o uso de datos públicos gubernamentales, con el fin de fomentar la utilización de datasets y, además, tener en cuenta algunos criterios de calidad del tipo cuantitativo, como ser: número de solicitudes de apertura de datos, número de dependencias/organismos estatales que abren datos, porcentaje del conjunto de datos abiertos estratégicos que fueron publicados, número de conjuntos de datos descargados versus número de conjuntos de datos publicados, o bien el número de conjuntos de datos visitados versus los datasets descargados. Otros trabajos, apuntan a los criterios de calidad orientados a la interoperabilidad, en donde proporcionar orientaciones y buenas prácticas para el desarrollo de estrategias de apertura de datos que cumplan con los principales estándares de calidad e interoperabilidad de datos abiertos de los países mejor calificados en índices internacionales



como el Open Data Barometer [ODB21] y el Global Open Data Index [AOK21]. Además, en las guías del dominio de información del Marco de Referencia del Ministerio de Tecnologías de la Información y las Comunicaciones (MinTIC), el modelo de calidad de datos ISO 25012, la Carta Internacional de datos abiertos y el Marco de Interoperabilidad para el Gobierno Digital [GES21].

Otros autores más recientes [CAD19], [VEL19], [LEO19], [SCH12], presentan análisis del estado actual en el ámbito de los datos abiertos, así como de los estándares internacionales y buenas prácticas de calidad de datos con el fin de proponer un marco de referencia que posibilite la publicación de datos abiertos con un nivel de calidad adecuado. Otros proyectos [BAR20], se orientan a la calidad, a través del análisis de la información publicada en los geoportales, con el objeto de medir el grado de reutilización de sus conjuntos de datos geoespaciales, esto es, debido a que no se cuenta con estándares específicos de análisis de calidad. Además, otros trabajos [ROY19] indican que la mayoría de los portales parecen funcionar como meros repositorios de datos, descuidando aquellos aspectos que promueven el uso de datos por parte del público no experto, la rendición de cuentas, la participación ciudadana y la calidad de los datos que se divulgan (por ejemplo, definición de los metadatos utilizados).

3.5. Medición de calidad

A modo de mediciones en aspectos de calidad, existen medidas o índices en diferentes gobiernos, por ejemplo, el Barómetro de Datos Abiertos, ODB [ODB21] de la World Wide Web Foundation, es una medida global de cómo los gobiernos publican y utilizan los datos abiertos para la rendición de cuentas, la innovación y el impacto social.

Otras mediciones fueron desarrolladas por The Global Open Data Index [AOK21], GODI [AOK21], que es el punto de referencia mundial anual para la publicación de datos gubernamentales abiertos, gestionado por la Open Knowledge Foundation [OKF20]. Funciona como una encuesta de crowdsourcing, que mide la apertura de los datos gubernamentales a través de la metodología GODI [MET17]. Otro proyecto es el Open Government Data de la Organisation for Economic Cooperation and Development, OECD [OEC20], su objetivo es avanzar en la evaluación de impacto de Open Government Data (OGD), para este caso, su índice evalúa los esfuerzos de los gobiernos para implementar datos abiertos en las tres áreas críticas: apertura, utilidad y reutilización de los datos gubernamentales; otras medidas son las puntuaciones otorgadas a la calidad de los datos abiertos, este es el caso del esquema modelo de cinco estrellas de Berners-Lee [OPD12], o los informes sobre el estado de los datos de gobierno abierto elaborados por fundaciones comprometidas con el impacto de estos datos [DSF21]. Adicionalmente, se llevan a cabo iniciativas de Datos Abiertos para Políticas Públicas en América Latina con potencial para mejorar la calidad de las políticas públicas en la región [ODD21] [OD421] financiadas por la CEPAL [CEP21] y el Consorcio World Wide Web en Brasil, W3C Brasil [W3B21].

Aunque varias organizaciones hacen un muy buen trabajo en el tratamiento de los aspectos de calidad en los datos abiertos, existen otros enfoques como los que propone el Gobierno de España a través de su manual práctico para mejorar la calidad de los datos abiertos [GOE17]. También hay otros puntos a considerar en cuanto a calidad, como los Principios de la Open Data Charter, Carta Internacional de Datos Abiertos [ODC21], que tiene los seis principios que fueron desarrollados en 2015 por gobiernos, sociedad civil y expertos de todo el mundo. Presenta un conjunto de estándares establecidos a nivel mundial sobre cómo publicar datos. Por otro lado, la



Norma ISO/IEC 25012 [ISO08], que especifica un modelo general de calidad de datos que se define en un formato estructurado dentro de un sistema informático, se utiliza ampliamente en este contexto. Por otro lado, el Estándar Universal de Calidad de Datos [CAI15] establece los criterios que debe cumplir un conjunto de datos para ser considerado de calidad e interoperable. Estos criterios están definidos por el estándar universal, con las dimensiones de calidad de datos (CDDQ) propuestas por Dan Myers en [DIM18].

3.6. Herramientas de validación de datos abiertos

En esta sección se analizan algunos de los trabajos relacionados con respecto a las herramientas de validación para analizar la calidad de los datasets de datos abiertos gubernamentales.

En el trabajo realizado por [MMH17], se presenta una propuesta de un prototipo en el que su diseño funcional plantea que se creen y se validen, distintas reglas a través del ingreso de un dataset. Esta aplicación permite analizar métricas de consistencia lógica para evaluar un conjunto de datos abiertos, como ser su longitud máxima o mínima definida, tipo de datos o bien dominio de valores (conjunto de valores posibles). A través de este prototipo se visualizan los registros de datos abiertos que no cumplen con las métricas de consistencia lógica y así obtener el índice de calidad según cada una de éstas. En la Figura 3.1 se puede observar una pantalla del prototipo que contiene las opciones para obtener el dominio de los campos del dataset, los rangos y la función principal de la generación de reglas de consistencia para el conjunto de datos abiertos ingresado.

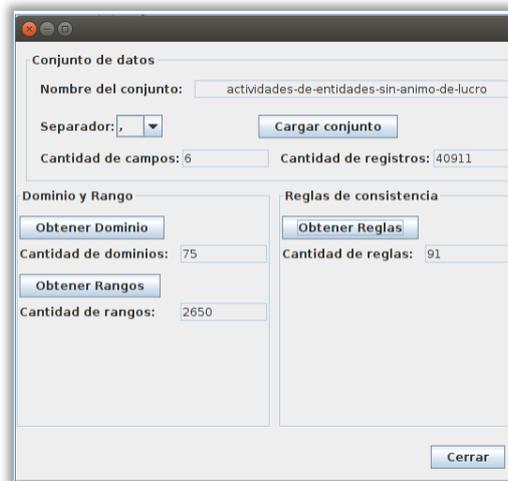


Figura 3.1 – Prototipo QILOGD. Pantalla de vista de conjunto de datos.

Otro de los trabajos realizados [AAD19], propone un prototipo de software para la evaluación parcial de los principios internacionales de datos abiertos, a través del análisis de un caso de estudio. Esta investigación colaboró con la validación del cumplimiento de la estrategia de gobierno para con los principios establecidos por el Open Government Data [OEC20]. Este prototipo analiza la información extraída de un dataset, la cual se realiza a través de una API y en la que debe llevar a cabo la generación de la estructura de los datos extraídos. Esta generación de estructura se efectúa en forma manual a través de una interfaz en la aplicación propuesta. Una vez concluido con estos pasos, posteriormente, se almacenan los datos en una base de datos (MongoDB), y su análisis se basa en distintas métricas de calidad de datos como ser: si se



encuentran completos, accesibles, actualizados, etc. En la Figura 3.2 se muestra la pantalla del prototipo de este trabajo, en el que se solicitan los valores correspondientes a la carga de la estructura del datasets tratado.

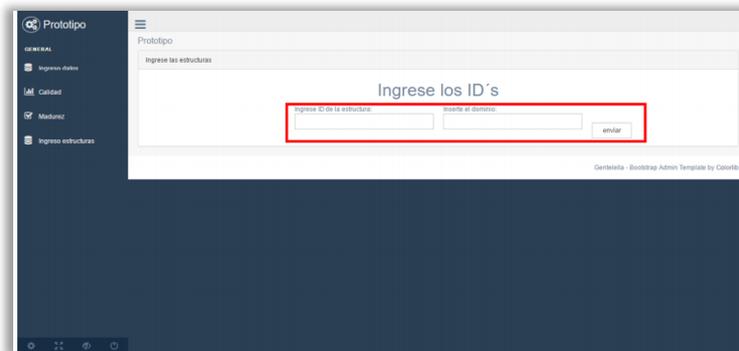


Figura 3.2 – Prototipo. Pantalla de ingreso de estructura de datos.

En la Figura 3.3. se visualiza la pantalla de validación de calidad de datos abiertos, como se puede observar, se muestra una grilla con los resultados obtenidos de la lógica de evaluación de principios de calidad de datos abiertos.

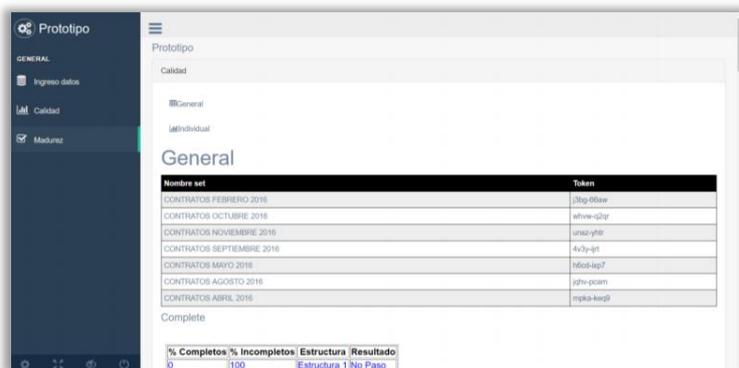


Figura 3.3 – Prototipo. Pantalla de validación de calidad de datos abiertos.

Otro de los trabajos desarrollados para la validación de calidad [DIX20], propone la ampliación de una herramienta de código abierto para estudiar la calidad de los datos abiertos orientados a datos de la salud. Este estudio fue desarrollado por OHDSI, Observational Health Data Sciences and Informatics [OHD21], que es una colaboración interdisciplinaria de múltiples partes interesadas para crear valor con los datos de salud a través de análisis a gran escala. En esta propuesta puntual gracias a la utilización de APIs, se enfocó principalmente en los métodos de medición para calcular la integridad de los datos, análisis de variables de tiempo y la entropía de información, que es la tasa promedio de la información, la cual es producida por una fuente de datos. En la Figura 3.4, se puede observar la ampliación de la herramienta propuesta, junto con su análisis de datos que fueron obtenidos de distintas fuentes de datos. Esto no sólo permite analizar datos, sino que, además permite colaborar con futuros proyectos sobre desafíos de calidad de datos.

Otro prototipo de herramienta que permite analizar la calidad de datos abiertos [MED21], es un proyecto que se llevó a cabo para evaluar la utilidad de conjuntos de datos específicos. Esta



propuesta consiste en, un módulo escrito en Python para programadores y una aplicación escrita en lenguaje R, cuya interfaz permite que cualquier usuario pueda realizar consultas. En la Figura 3.5, se muestra el primer prototipo de la aplicación que incluye algunos de los pilares de calidad de datos abiertos. Como se puede observar, se realiza un análisis en base a las métricas de calidad de los datos abiertos analizados. Algunos de los conceptos tratados con la aplicación son: completitud, veracidad, consistencia, metadatos y valores únicos entre otros. Básicamente, este prototipo enfoca el cálculo de métricas objetivas para evaluar la calidad de los datasets seleccionados.

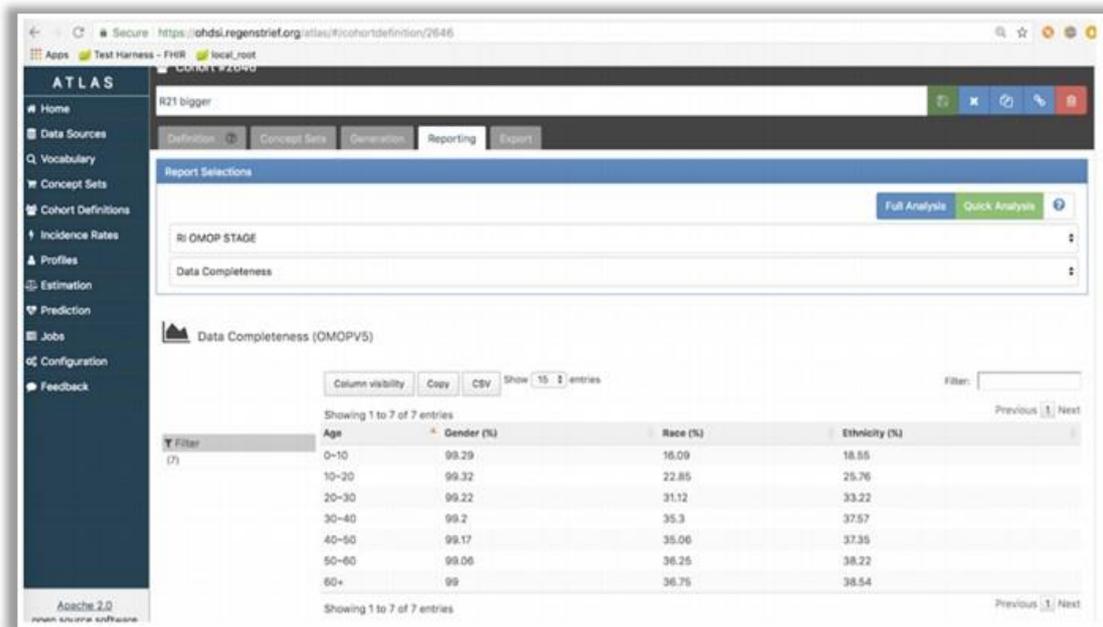


Figura 3.4 – Pantalla de la herramienta OHDSI ATLAS.

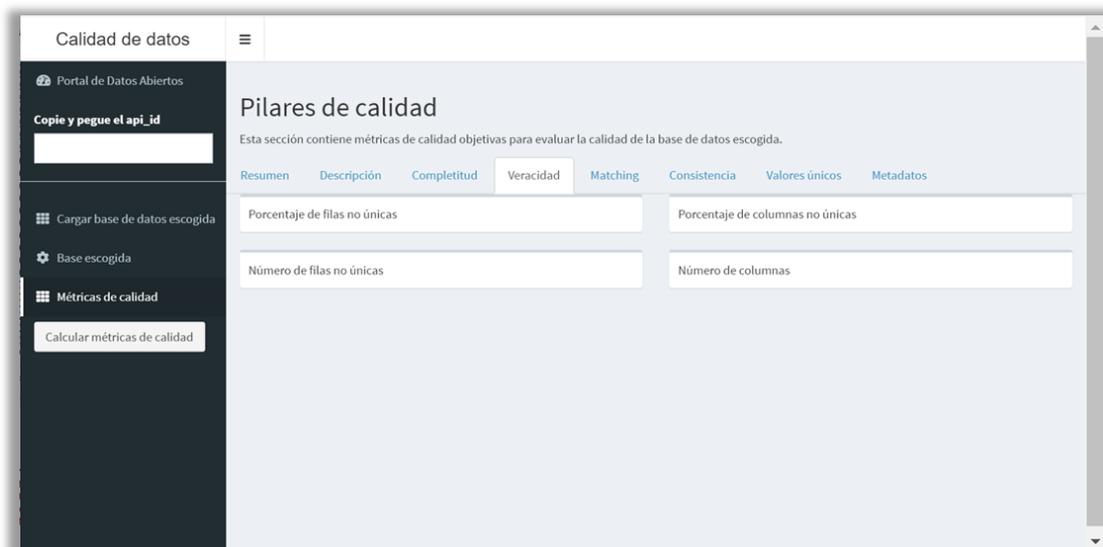


Figura 3.5 – Pantalla del Prototipo para medir la calidad de datos con métricas.



Otro trabajo interesante a nivel validación de calidad de datos, es el realizado por la Comisión Europea, INSPIRAR GEOPORTAL [CEI21], quienes trabajaron en un validador de metadatos sobre datos espaciales europeos con el objetivo que se cumpla con los reglamentos establecidos por normativas europeas (Directiva INSPIRE) [GEC21]. Este registro obtiene datos del conjunto de información geográfica oficial junto con la identificación de cada serie de éstos, la norma técnica de aplicación, sus parámetros de calidad, fechas de referencia y el organismo responsable de su producción y mantenimiento. Esta herramienta de validación trabaja con los metadatos que se analizan y validan si cumplen o no con las normativas y los esquemas. En la Figura 3.6, se muestran los validadores que indicarán si existen o no errores en los archivos, y además se visualiza una leyenda de cómo corregirlos y se incluyen buenas prácticas. Este validador se encuentra en su sitio web [VAL21]. En la Figura 3.7, se muestran los errores encontrados, luego del proceso de verificación de calidad de los metadatos. Como se puede observar, se identifican que tipo de error es el que fue detectado.

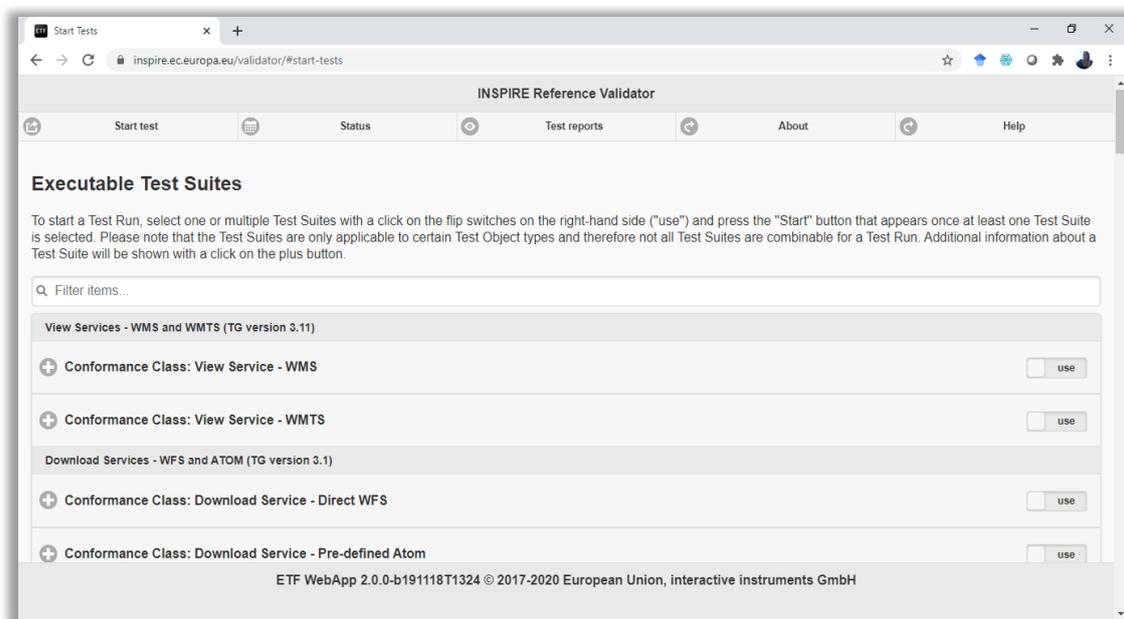


Figura 3.6 – Pantalla de la herramienta de validación de metadatos INSPIRE.



Figura 3.7 – Una vez completado el proceso de validación, se muestran los errores encontrados.



Por otra parte, la Fundación de Conocimiento Abierto, Open Knowledge Foundation [OKF20], mantiene un seguimiento sobre el estado de los datos abiertos del gobierno de varios países de todo el mundo, casi 100 países. Su trabajo se centra en el análisis y la medición de distintos conjuntos de datos abiertos gubernamentales por país [AOK21]. Para cada conjunto de datos abiertos, se realiza una medición sobre: si los datos poseen licencia abierta, si tienen formatos abiertos en su totalidad, si son: legibles por algoritmos de software, descargables desde cualquier ubicación, actualizados, disponibles públicamente y si son gratuitos. En la Figura 3.8, se muestra el sitio web de Open Knowledge Foundation [OKF20], en el que se refleja el análisis de los datasets de Argentina con su correspondiente porcentaje de apertura en base a los criterios estudiados.

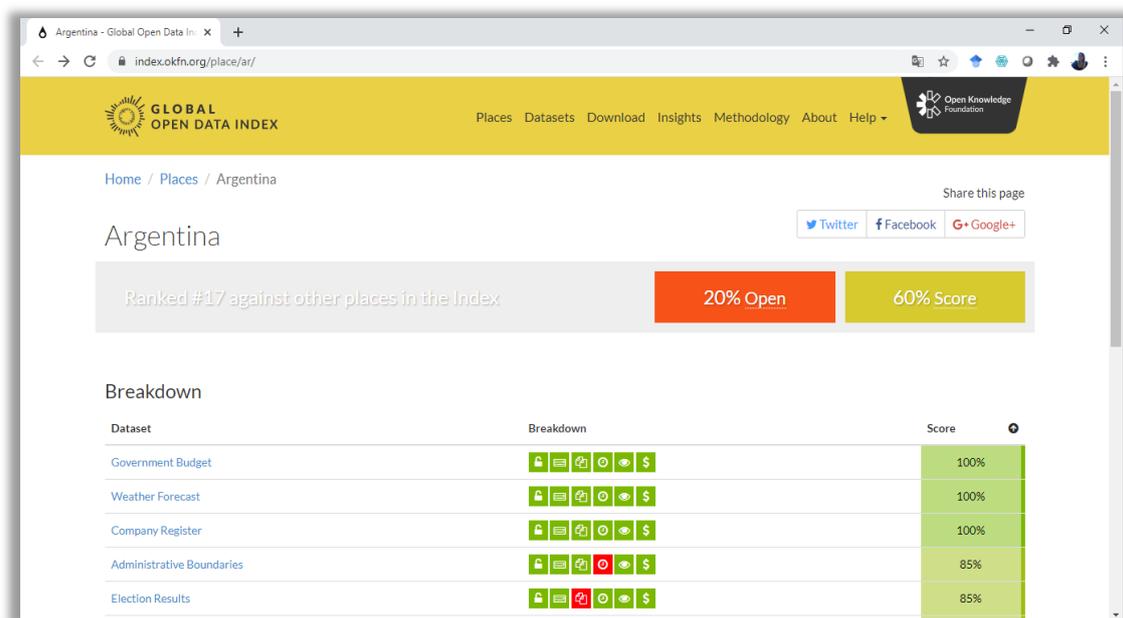


Figura 3.8 – Validación de calidad de datos abiertos de Argentina, con nivel de apertura.

Otro de los prototipos realizados [LBN17], propone un software para la evaluación de la calidad de datos abiertos. El proceso inicia con la extracción de datos mediante un cliente utilizando JAVA como lenguaje de programación. Esta información se almacena en una base de datos no relacional, MongoDB. Luego, se convierten los datos de JSON a una estructura que sea interpretada por el software RapidMiner, la cual permite diseñar procesos de minería de datos cajas que representan módulos o actividades llamados procesos, esto con el fin de realizar flujos de datos o control sin programación. Una vez que están todas las filas y columnas en RapidMiner se procede a obtener todos los datos necesarios para la aplicación de la fórmula de la métrica que se desea analizar, Para este trabajo se estudió la métrica de Trazabilidad, Completitud y Conformidad de los datos abiertos.

Otro de los trabajos es el de la organización Open Contracting Partnership [OCP21] alojada en Washington, D.C., USA, que propone una herramienta [OCD21] de validación de calidad de datos para evaluar los datos públicos de contrataciones abiertas. Estas validaciones se enfocan en una serie de elementos, como ser: integridad de datos, estructura de contratos predefinida y demás. En la Figura 3.9, se muestra el estándar de datos de contrataciones abiertas (OCDS, Open Contracting Data Standard), la cual se utiliza para formalizar cómo se pueden publicar los datos y documentos sobre contrataciones de una manera accesible, estructurada y repetible.

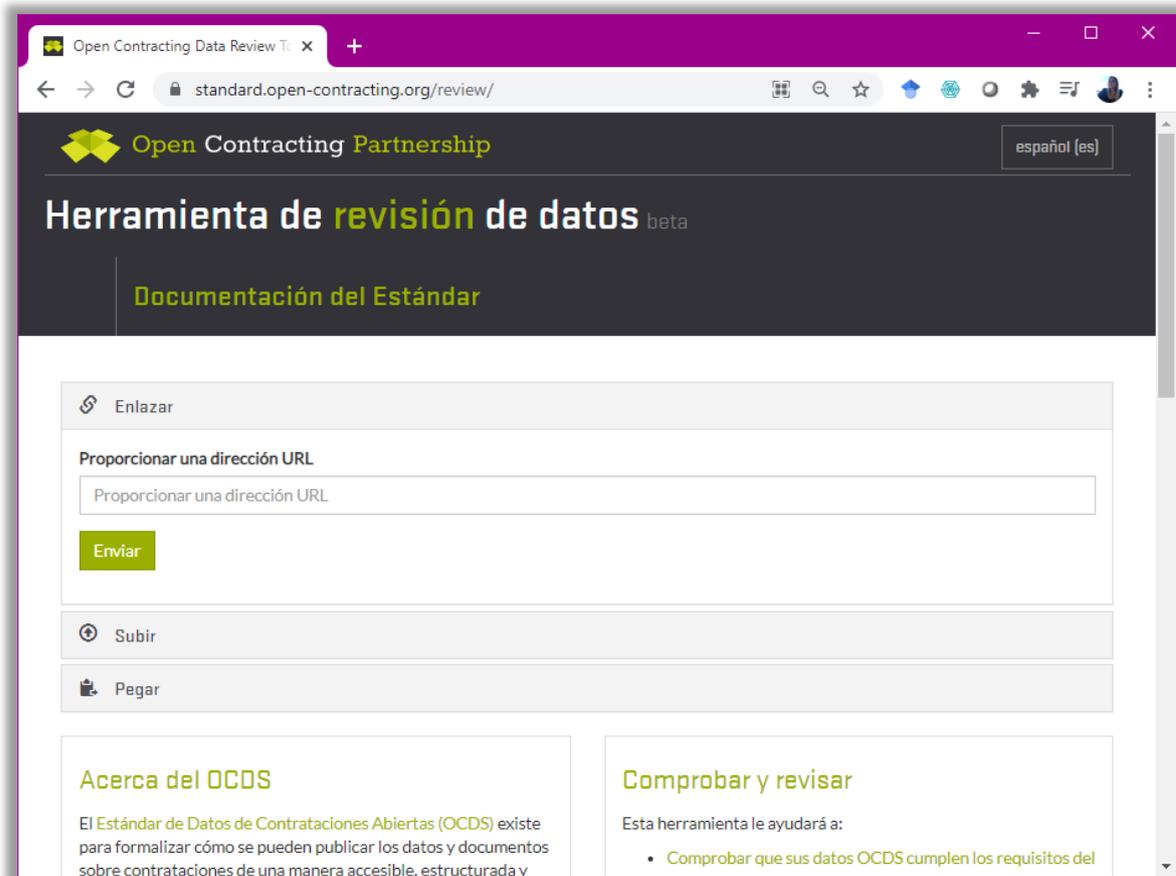


Figura 3.9 – Herramienta de revisión de datos beta OCDS.

Para esta tesis, no se toman en cuenta las herramientas comerciales de validación de datos genéricos (DataCleaner [DCL20], Data Quality Services [MIC21], Open Studio for Data Quality de Talend [TAL21], Oracle Enterprise Data Quality [ORA21], entre otros), sino que se apunta a validadores de datos abiertos de desarrollo no comercial.

3.7. Resumen de Trabajos relevados

Estos trabajos fueron analizados para identificar métricas que puedan medir y evaluar aspectos sobre los archivos de datos abiertos. Algunos de estos trabajos proponen estándares de medición y buenas prácticas para la evaluación de los conjuntos de datos disponibles en portales de datos abiertos. Estos estudios consideran algunas cuestiones orientadas a la medición de archivos en sitios web públicos y publican estadísticas sobre el número de archivos descargados, número de conjuntos de datos, licencias, metadatos o criterios de calidad orientados a la interoperabilidad de software y formato de archivo, pero en ninguno de estos casos, se hace un estudio del contenido de los conjuntos de datos gubernamentales, es decir, qué valores tienen y en qué estado se encuentran. Por otro lado, se presentaron varios autores que proponen estándares de publicación para la apertura de archivos y visualización de contenidos para el ciudadano, pero no un análisis detallado del contenido de conjuntos de datos abiertos. Otros trabajos de investigación



están enfocados a dar a conocer este nuevo paradigma, por lo que proporcionan un repositorio de ubicaciones geográficas de países con portales de datos abiertos en todo el mundo [ODI16], [POR11], pero son solo accesos directos y no presentan un análisis de los conjuntos de datos.

En cuanto a los criterios de medición, algunos organismos internacionales, Global Open Data Index [AOK21], presentan índices que surgen del análisis de un conjunto de aspectos para cada país, por ejemplo, se muestra que solo el 11% de los conjuntos de datos a nivel mundial son abiertos. Si bien este análisis es interesante, solo se estudian los puntos referentes a las licencias, si los archivos son legibles por una máquina, si se pueden descargar del portal oficial, si están actualizados, entre otros, pero esto no es un análisis desde el punto de vista del contenido que tienen los conjuntos de datos. Otros estudios sobre el impacto de las iniciativas de datos abiertos son: el Barómetro de Datos Abiertos [ODB21] y el Inventario de Datos Abiertos (ODIN) (Open Data Watch [ODW20], estos trabajos son casos en los que se observó que evalúan la cobertura y apertura de datos para continuar con políticas de datos abiertos, pero tampoco realiza un análisis interno del contenido de los conjuntos de datos, sino que su estudio está orientado a la disponibilidad de la estructura general de los archivos. Desde el enfoque de modelos de evaluación, existen organismos internacionales como la Carta Internacional de Datos Abiertos [XHA20] y [PIN04] que proponen como metodología de evaluación, algunos parámetros que pueden ser evaluados sobre el contenido, accesibilidad, funcionalidad, navegabilidad, actualizado y diseñado. Asimismo, es importante que la evaluación esté orientada al uso de una guía de buenas prácticas estándar [IND20] para facilitar la interoperabilidad y accesibilidad [PAS18] para mantener los principios de apertura. Desde el aspecto de calidad del contenido del dataset, existen estudios realizados en Argentina, Brasil y Paraguay [CAT20] que miden la calidad en función de su contenido, por ejemplo, datos incompletos, obsoletos e inválidos, entre otros. Esto nos lleva a pensar en otro tipo de abordaje de los temas de calidad, ya que una visión de calidad no se presenta desde el archivo, sino desde el contenido del archivo.

Esta tesis enfoca los aspectos que deben ser abordados y conciliados entre diversos sectores de la administración pública en este paradigma de transparencia gubernamental, la implementación de estándares de calidad en datos abiertos favorecerá a diversas organizaciones estatales que no solo brindan datos públicos a los ciudadanos, sino también a otras entidades estatales en todo el mundo. Con base en los estudios realizados, se detectó que existen problemas importantes, por un lado, se pudo observar que existen varios inconvenientes en las estructuras de los conjuntos de datos disponibles en los portales de datos abiertos y que los datos tienen varias deficiencias, por ejemplo, datos incompletos o vacíos en el interior o problemas de estructuras en las que el número de columnas y otros no está delimitado correctamente, el otro gran inconveniente es que no existen herramientas de control o validación para los conjuntos de datos, debido a que hay pocas investigaciones que centran su estudio en el contenido y la calidad de los datos aportados.

Con base en los hallazgos de este estudio, es posible afirmar que las herramientas de validación y análisis de métricas de calidad de datos abiertos son necesarias para mantener la validez e integridad del contenido. Estas herramientas favorecen a una alta posibilidad de obtener un análisis confiable sobre un determinado contexto. Esto permite obtener un estudio sobre los puntos a considerar y tener un “estado de salud” de los conjuntos de datos, que podría mejorarse, en caso de detección de fallas, por ejemplo, redundancia, falta de campos, falta de nombres que identifican lógicamente cada campo en el conjunto de datos, entre otros. Una herramienta de detección permite utilizar y reflejar la veracidad del contenido en un estudio con valor añadido para la ciudadanía. Por lo que, sobre los temas y trabajos relacionados con dicha temática, se presenta a continuación, el Capítulo 4 con que presenta una propuesta de métricas de calidad.



Capítulo 4 – Definición de Métricas para el análisis de Datasets Gubernamentales

Actualmente existen diversos conjuntos de datos (datasets) que son brindados por organismos gubernamentales, tanto a nivel nacional como internacional en sus portales de datos abiertos. Muchas veces, estos datos no poseen calidad, es decir, no están en un formato abierto, existen caracteres inválidos, o bien se encuentran campos incompletos entre otros problemas. La propuesta de esta tesis es generar métricas que permitan analizar la calidad de estos.

Los datos son un recurso indispensable para cualquier actividad de gestión pública, por lo que es necesario mantenerlos actualizados y completos. Si bien se puede determinar que cada vez son más los gobiernos que se embarcan en este nuevo paradigma de concepto de datos abiertos, todavía falta un largo camino por recorrer en esta temática, como se trató en el capítulo de trabajos relacionados.

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. “La reutilización de datos abiertos permite el desarrollo de nuevos productos y servicios digitales, creando oportunidades de desarrollo social y económico. Sin embargo, la reutilización de estos datos se enfrenta con diversas barreras en su expansión, debido a diversos problemas relacionados con la calidad de datos que van desde la incompletitud de los datos hasta la de actualización de estos. En concreto, un criterio relevante de calidad de datos abiertos es la comprensibilidad, ya que un proceso de interpretación errónea provocará ambigüedades o malentendidos que desmotivará su reutilización” [CAD19].

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, esta tesis se orienta en algunos aspectos más relevantes en cuestiones de calidad de datos sobre criterios estándares de la Norma ISO/IEC 25012 [ISO08], estándares universales de calidad de datos [CAI15], dimensiones de la calidad de los datos [DIM18], trabajos relevados y estudios realizados [MAR20] en esta temática.



La utilización de métricas de calidad favorece al encuadre de indicadores que permitan obtener un dato más limpio para facilitar el análisis final. “Los indicadores para calidad de datos son herramientas importantes que debemos tomar en cuenta en nuestros procesos de análisis ya que nos permite medir y controlar la eficiencia de nuestros procesos que derivarán en análisis y toma de decisiones dentro de una estructura organizacional” [GRA21].

Esta propuesta de métricas de calidad se basan en el estudio de guías y buenas prácticas de publicaciones de datos abiertos gubernamentales [TIP21], en el formato recomendado para los tipos de datos que está mayormente basado en las especificaciones de la W3C [W3C15], en la experiencia de estudios realizados con datasets relevados de diversos sitios gubernamentales de portales de datos abiertos [MAR20], varios criterios enfocados en la calidad del dato (Normas y estándares) y guías elaboradas por el Gobierno de la Ciudad de Buenos Aires, las cuales plantean una serie de criterios para tener en cuenta en la apertura y tratamiento del contenido de los datasets públicos. Estas [SEC21] fueron elaboradas por el equipo de la Dirección Nacional de Datos e Información Pública junto con la Iniciativa Latinoamericana por los Datos Abiertos (ILDA), y su implementación fue posible gracias al financiamiento del fondo de transparencia del BID, que se beneficia de las generosas contribuciones de los gobiernos de Canadá, Noruega y Suecia [EST21].

En las siguientes secciones se presentan las métricas propuestas para el tratamiento de la calidad de los conjuntos de datos en formatos abiertos a raíz de un trabajo de investigación propio y publicado en la IEEE [MAR21]. Las métricas se clasifican de la siguiente manera:

- **Métricas Críticas:** 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.
- **Métricas No Críticas:** 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.

4.1. Métricas Críticas

A continuación, se presentan cada una de las métricas propuestas en esta categoría.

4.1.1. Números Decimales

Esta métrica analiza los datos del tipo decimal, verificando si cumplen con las siguientes recomendaciones:

- Valores numéricos con 2 dígitos decimales;
- Formato de separación de decimal con “.” (punto);
- Valores numéricos sin separador de mil.



Como información adicional, se podría considerar la cantidad de columnas con valores numéricos decimales mal cargados en un dataset verificado. Además, este aspecto permite obtener un análisis estimativo de los tipos de datos almacenados en los campos, siendo éstos del tipo: decimal, texto, numérico (entero) y/o coordenada (se consideran los campos con valor decimal y con título de columna: x, y, lat, long, latitud, longitud, _latitud o _longitud). Uno de los objetivos fundamentales de los datos abiertos gubernamentales, es la interoperabilidad entre softwares, por lo que es necesaria la identificación del tipo de dato para un correcto intercambio de datos entre diversos programas.

Esta tesis se aboca a una propuesta de métrica para uno de los tipos de datos, siendo este el de números decimales, por lo que se relevaron algunas de las recomendaciones que fueron agregadas a raíz del relevamiento para la publicación de datos abiertos en este contexto. Éstas son:

- No deberá utilizarse espacios en blanco en los valores de datos;
- Para los números negativos se debe incluir el símbolo menos (-), antes del número, sin espacio en blanco intermedio. En caso de que tenga espacio, se considerará como tipo texto.

En la Figura 4.1, se muestra un ejemplo de un dataset extraído del portal de datos abiertos de la Ciudad de Buenos Aires [BAC21] llamado “*espacios-verdes-por-habitante-por-comuna.csv*”, éste tiene casos que no cumplen con el formato de números decimales explicado anteriormente. Allí se muestran algunos registros que poseen más de un decimal en la columna “superficie_por_habitante”.

```
1 comuna,año,superficie_por_habitante
2 Total,2011,0.628
3 1,2011,2.346
4 2,2011,0.277
5 3,2011,0.040
6 4,2011,0.385
7 5,2011,0.018
8 6,2011,0.143
9 7,2011,0.177
10 8,2011,2.287
11 9,2011,0.682
12 10,2011,0.129
13 11,2011,0.308
14 12,2011,0.799
15 13,2011,0.655
16 14,2011,0.970
17 15,2011,0.128
18 Total,2012,0.624
19 1,2012,2.332
20 2,2012,0.289
21 3,2012,0.037
22 4,2012,0.378
23 5,2012,0.018
```

Figura 4.1 – Conjunto de datos para el análisis de Números Decimales.



Se observa que existen registros que no cumplen con los criterios de buenas prácticas explicadas anteriormente. Por ejemplo, se visualiza que el registro número 2, que tiene un valor de “0.628”, no cumple con el criterio del tipo cantidad de decimales ya que es mayor a 2. Esto mismo sucede con los otros registros visualizados en la Figura 4.1, debido a que todos tienen más de 2 dígitos decimales en su tercera columna.

4.1.2. Registros Duplicados

Esta métrica propuesta permite detectar y mostrar los registros que se encuentran duplicados en el conjunto de datos analizado.

Uno de los puntos fundamentales para tener en cuenta en las fuentes de datos abiertos gubernamentales, es la correcta utilización de estos, es por ello por lo que detectar los casos de duplicación de registros favorece en una mejor manipulación de estos. Esto a su vez, permite definir estructuras de datos y simplicidad en el tratamiento de los distintos procesos que se podrían utilizar con ellos, por ejemplo, procesos de Extracción, Transformación y Carga, ETL (Extract, Transform and Load) que brindan a las organizaciones una gestión de datos desde múltiples fuentes.

Es necesario tener una fuente de datos de calidad, para llevar a cabo un modelo flexible y consistente, es decir, sin contar con datos sucios, ya que estas fuentes serán utilizadas por diversos organismos estatales como así también ciudadanos, que podrían utilizarlas para enfoques predictivos de casos. Como menciona el autor Beltrán [BEL14] en relación con los datos sucios, indica que “los datos suelen tener problemas que afectan significativamente a la precisión del modelo predictivo y pueden llevar a conclusiones erróneas, por ejemplo, pueden no contener suficiente información, pueden no estar elegidos al azar o, simplemente, pueden ser erróneos”. Algunos de los proyectos en gestión pública [GOB17], [GOM18], procesan datos abiertos para detectar variables que predigan incumplimientos, o bien se desarrollaron modelos estructurales, proyectos de simulaciones, o bien se utilizan los datos como parte de un pronóstico para relevar inspecciones de manera más eficaz.

Uno de los enfoques que garantiza la calidad en los datos, se relaciona con el aspecto de dimensiones, para esta métrica se propone la dimensión de Unicidad, esta se encuentra en el factor de calidad: No-duplicación en el cual se estudia el grado de duplicación de la fuente de datos analizada. Dentro de esta, se trabaja en varios aspectos, por ejemplo, para esta métrica puntual, en la redundancia de los registros de un conjunto de datos.

La dimensión de calidad de datos, unicidad [CAL12], indica el nivel de duplicación de los datos. Esto ocurre cuando un objeto del mundo real se encuentra representado más de una vez en los datos, esto es, varias tuplas (registros) representan exactamente el mismo objeto. Por lo que la duplicación es cuando la misma entidad aparece repetida de manera exacta. Básicamente, la unicidad [MMH17], mide el grado en que un dato está libre de redundancias en amplitud, profundidad y alcance. En amplitud las propiedades y clases representadas, En alcance, una base de conocimiento en donde múltiples ejemplares representan el mismo objeto; En profundidad donde múltiples valores de una propiedad son únicos. Es decir que, “las medidas de unicidad permiten entender que los valores distintos de un elemento de datos aparecerán una sola vez reflejados en el conjunto de datos” [GRA21].



La correcta “detección de registros duplicados en archivos electrónicos que contienen información acerca de objetos del mundo real [...], es un proceso de importancia fundamental si deseamos generar, analizar y diseminar información estadística de buena calidad. En particular, para encuestas de establecimientos comerciales, es de suma importancia, sobre todo en las primeras fases, detectar oportunamente registros duplicados si deseamos construir o actualizar de manera correcta un marco muestral de establecimientos: los elementos duplicados dentro de un marco muestral tienen efectos indeseables, como sobreestimación de totales poblacionales, o generación de muestras sesgadas para la realización de nuevas encuestas” [ALB11].

En las Figuras 4.2 y 4.3 se muestra un ejemplo de registros con casos de duplicados (métrica de redundancia), estos son producto del análisis efectuado del dataset que fue extraído del portal de datos abiertos de la Ciudad de Buenos Aires [BAC21], llamado “*padron-de-operadores-organicos-certificados.csv*”. Para ello se identifica la cantidad de casos detectados con duplicación de registros, indicando la cantidad afectada sobre el total de registros del dataset.

```
1 pais_id,pais,provincia_id,provincia,departamento,localidad,rubro,productos,categoria_id,categoria_desc,Certificadora_id,certificadora_denominacion,razón social,establecimiento
2 32,ARGENTINA,62.4599,RIO NEGRO,VALLE AZUL,INDEFINIDA,EMPAQUE Y FRIO,MANZANAS Y PERAS,2,Elaboradores,3,Letis S.A.,SARALEGUY SANDRO SERGIO,NC
3 32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,AGRICULTURA,SOJA,1,Productores,2,OIA S.A.,ADD CROPS S.A.,BARADERO 1
4 32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,AGRICULTURA,"GIRASOL, MIJO, RESIDUO VEGETAL",1,Productores,2,OIA S.A.,ADD CROPS S.A.,BARADERO 2
5 32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,AGRICULTURA,"MAIZ, AVENA, CAMPO NATURAL",1,Productores,2,OIA S.A.,ADD CROPS S.A.,EL BUEN RETIRO
6 32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO
7 32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE, TRIGO",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL HORIZONTE
8 32,ARGENTINA,6,BUENOS AIRES,CARMEN DE PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO ROSSI S.H.,DON GISLEMO
9 32,ARGENTINA,6,BUENOS AIRES,CARMEN DE PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO ROSSI S.H.,ONELIA
10 32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,FRUTICULTURA/HORTICULTURA,"NUEZ PECAN, PASTURAS, INCULTO",1,Productores,2,OIA S.A.,ALVAREZ ANA MARIA,EL PAMPA ORGANICO
11 32,ARGENTINA,6,BUENOS AIRES,VILLARINO,INDEFINIDA,GANADERIA,GANADERIA BOVINA,1,Productores,1,Argencert S.R.L.,ANIJG AGENCY S.A.,EL FORTIN
12 32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO
13 32,ARGENTINA,6,BUENOS AIRES,MERCEDES,INDEFINIDA,FRUTICULTURA/HORTICULTURA,ARANDANO,1,Productores,2,OIA S.A.,ARBE JUAN LUCAS ,NATURAL PAMPA
14 32,ARGENTINA,6,BUENOS AIRES,LOS CARDALES,INDEFINIDA,FRUTICULTURA/HORTICULTURA,"ARANDANO, FRAMBUESA, MORA, TOMATE, PIMIENTO, ZAPALLO, INCULTO",1,Productores,2,OIA S.A.,ARGENTINA SPECIALITY S.R.L.,LA MORENITA
15 32,ARGENTINA,6,BUENOS AIRES,SAN PEDRO,INDEFINIDA,FRUTICULTURA/HORTICULTURA,"ARANDANO, DESCANSO",1,Productores,2,OIA S.A.,AUTOSOL PARTS S.A.,SAN JOSE
16 32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,TRIGO Y SOJA,1,Productores,3,Letis S.A.,BANDERITAS S.A.,JUAN ALBERTO
17 32,ARGENTINA,6,BUENOS AIRES,BENITO JUAREZ,INDEFINIDA,AGRICULTURA,LINO DORADO Y SOJA,1,Productores,3,Letis S.A.,BANDERITAS S.A.,SANTA RITA
18 32,ARGENTINA,6,BUENOS AIRES,BENITO JUAREZ,INDEFINIDA,AGRICULTURA,TRIGO Y SOJA,1,Productores,3,Letis S.A.,BANDERITAS S.A.,LA COLORADA
```

Figura 4.2 – Dataset con Registros Duplicados (Registros 1 al 17).

En la Figura 4.2, se visualiza que el contenido del registro número 6 se encuentra duplicado en la fila y/o número de registro 12, bajo el detalle del registro:

"32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO".



ID	Record Text
1101	32, ARGENTINA, 46, LA RIOJA, AIMOGASTA, INDEFINIDO, PROCESAMIENTO CULTIVOS INDUSTRIALES, ACEITE DE JOJOBA, 2, Elaboradores, 2, OIA S.A., FUERTE DEL BAÑADO S.A., NC
1102	32, ARGENTINA, 46, LA RIOJA, VILLA UNION, INDEFINIDO, SECADERO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., "GARROT, CARLOS PEDRO", NC
1103	32, ARGENTINA, 46, LA RIOJA, AIMOGASTA, INDEFINIDO, PROCESAMIENTO CULTIVOS INDUSTRIALES, "ACEITE DE JOJOBA, ACEITE DE CHIA", 2, Elaboradores, 2, OIA S.A., GSP EMPRENDIMIENTOS AGROPECUARIOS SA, NC
1104	32, ARGENTINA, 46, LA RIOJA, LA RIOJA, INDEFINIDO, PROCESAMIENTO FRUTALES Y HORTALIZAS, "ACEITE DE OLIVA, ACEITE DE OLIVA SABORIZADO", 2, Elaboradores, 2, OIA S.A., INDUSTRIAS QUIMICAS Y MINERAS TIMBO S.A., NC
1105	32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, ELABORACION DE PASAS DE UVA, PASAS DE UVA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC
1106	32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, SECADO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC
1107	32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, SECADO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC
1108	32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, SECADO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC
1109	32, ARGENTINA, 46, LA RIOJA, LA RIOJA, INDEFINIDO, EXTRACCION DE ACEITE, OLIVA, 2, Elaboradores, 1, Argencert S.R.L., OLIVOS ARGENTINOS SA, NC
1110	32, ARGENTINA, 46, LA RIOJA, CAMPO GRANDE, INDEFINIDO, BODEGA VITIVINICOLA, VINO, 2, Elaboradores, 3, Letis S.A., VALLE DE LA PUERTA S.A., NC
1111	32, ARGENTINA, 50, MENDOZA, MAIPU, CONQUIMBITO, PROCESAMIENTO FRUTALES Y HORTALIZAS, "PULPA DE MANZANA, PURE DE MANZANA, PULPA DE PERA, PURE DE PERA", 2, Elaboradores, 2, OIA S.A., ADETECH S.A., NC
1112	32, ARGENTINA, 50, MENDOZA, MAIPU, INDEFINIDO, BODEGA VITIVINICOLA, VINO, 2, Elaboradores, 1, Argencert S.R.L., ANDILLIAN SA, NC
1113	32, ARGENTINA, 50, MENDOZA, TUNUYAN, INDEFINIDO, PROCESAMIENTO FRUTALES Y HORTALIZAS, MANZANA, 2, Elaboradores, 2, OIA S.A., ARGENTBIO S.R.L., NC
1114	32, ARGENTINA, 50, MENDOZA, SAN CARLOS, INDEFINIDO, PROCESAMIENTO FRUTALES Y HORTALIZAS, "AJO, ZAPALLO", 2, Elaboradores, 2, OIA S.A., ARGENTBIO S.R.L., NC
1115	32, ARGENTINA, 50, MENDOZA, LUJAN DE CUYO, UGARTECHE, PROCESAMIENTO FRUTALES Y HORTALIZAS, "PULPA DE PERA, PULPA CONCENTRADA DE PERA, PULPA DE MANZANA, PULPA CONCENTRADA DE MANZANA, PULPA DE DURAZNO, PULPA CONCENTRADA DE DURAZNO, PULPA DE DAMASCO, PULPA CONCENTRADA DE DAMASCO, PULPA DE ZAPALLO, PULPA DE ZANAHORIA, PURE DE PERA CONCENTRADO, PURE DE PERA, PURE DE MANZANA, PURE DE MANZANA CONCENTRADO, PURE DE DURAZNO, PURE DE DURAZNO CONCENTRADO, PURE DE DAMASCO, PURE DE DAMASCO CONCENTRADO, PURE DE ZAPALLO, PURE DE ZANAHORIA, PURE DE BATATA", 2, Elaboradores, 2, OIA S.A., AVA S.A., NC
1116	32, ARGENTINA, 50, MENDOZA, GUAYMALLEN, INDEFINIDO, PROCESAMIENTO FRUTALES Y HORTALIZAS, AJO, 2, Elaboradores, 2, OIA S.A., BAMENEX S.A., NC
1117	32, ARGENTINA, 50, MENDOZA, GUAYMALLEN, INDEFINIDO, PROCESAMIENTO FRUTALES Y HORTALIZAS, AJO NEGRO, 2, Elaboradores, 2, OIA S.A., BAMENEX S.A., NC

Figura 4.3 – Dataset con Registros Duplicados (Registros 1101 al 1117).

Por otro lado, en la Figura 4.3, se detecta que el registro número 1106 se encuentra duplicado en más de un caso, como ser en el número de registro 1107 y 1108. A continuación, se muestra el detalle de registro:

"32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, SECADO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC".

4.1.3. Datos Faltantes y/o Completos

Esta métrica propuesta se analiza en dos partes. Como primera parte, permite detectar la cantidad de registros en los que todos sus campos están completos en un conjunto de datos, y el porcentaje que representa sobre el total de los registros del dataset. Como segunda parte, detecta y muestra los registros que se identifiquen bajo los siguientes casos:

- **Nulos:** sin Datos ni espacios en los campos.
- **Vacíos:** sin Datos y con espacios en los campos.
- **No Disponibles:** con datos que indiquen "N/D", "N/A", "NULL", "-", "--" o bien "---" en el conjunto de datos.

La importancia de tener todos los valores cagados en sus correspondientes campos permite un correcto análisis de manera completa sobre los datos, como así también lograr tener información para una adecuada interpretación, y así efectuar un estudio en detalle de lo analizado.



Básicamente, la falta de valores en los campos del dataset, proporciona una delgada línea a la confusión y/o error de interpretación de casos, ya que muchos de estos datos abiertos, son utilizados en tablas dinámicas, algoritmos estadísticos, historias de datos abiertos [BSA21], visualizaciones de gráficos o bien desarrollos de software.

Al igual que la métrica de detección de casos de los Registros Duplicados de la sección anterior, para el análisis de esta métrica, se toman en cuenta medidas cuantitativas de calidad de datos. Este enfoque que garantiza la calidad en los datos, se orienta el aspecto de dimensiones al concepto de: Completitud. “El nivel de completitud de datos refleja el grado en el que todos los atributos de un dato están presentes, lo que permite tener una visión clara sobre la integridad de los elementos a estudiar” [GRA21]. Esta dimensión de calidad de datos, completitud [CAL12], indica si el sistema de información contiene todos los datos de interés, y si los mismos cuentan con el alcance y profundidad que sea requerido. Además, indica que existen dos factores de la completitud: cobertura y densidad:

- La cobertura se refiere a la porción de datos de la realidad que se encuentran contenidos en el sistema de información.
- La densidad se refiere a la cantidad de información contenida y faltante acerca de las entidades del sistema de información.

En un modelo relacional, la completitud (en particular, la densidad) se caracteriza principalmente por los valores nulos, cuyo significado a pesar de ser variado, es importante conocer. Un nulo puede indicar que dicho valor no existe, que existe, pero no se conoce, o que no se sabe si existe en el mundo real. Para esta métrica se considera oportuno, estudiarla desde el factor de calidad: Densidad.

Uno de los temas importantes, para con esta dimensión, se debe a que uno de los mayores problemas en completitud, se da por valores nulos o vacíos. Esto conlleva a pensar que un valor puede reportarse como nulo, porque tal vez se omitió dicha información por parte de la fuente de datos, o bien por un error de almacenamiento en la base de datos y/o sistema que lo generó. En base a esto, es que se considera necesario, identificar estos casos para un análisis más detallado de éstos.

Desde la guía de sugerencias de buenas prácticas para la publicación de datos en formatos abiertos, se sugiere el tratamiento e identificación de los valores nulos, desconocidos o en blanco en un conjunto de datos, y que los elementos o celdas en blanco se interpretarán siempre como "valor ausente" [NUL21]. Por este motivo, esta tesis propone 3 grupos de casos para darle tratamiento a este tipo de aspectos: *Nulos*, *Vacíos* y *No Disponibles* (explicados anteriormente). Otra de las recomendaciones que se indican, es “no dejar celdas vacías en filas bajo la presunción de que valores en blanco posteriores a un valor positivo contienen implícitamente a ese mismo valor en una suerte de "agrupamiento conceptual". Este error es muy común en la construcción de planillas de cálculo y suele generar problemas graves cuando cambia el orden original de las filas. Otro tema para tener en cuenta es que impide el uso de tablas dinámicas y otras formas de analizar los datos” [CEL21], es decir, se sugiere indicar un valor redundante en caso de que corresponda relación lógica, antes que dejar un campo vacío. Esto es con el fin de facilitar el análisis de datos en un agrupamiento lógico conceptual. En la Figura 4.4 extraída de [CEL21], se muestra un ejemplo en el que se indica lo no recomendado y recomendado para los casos de celdas vacías relacionadas.



No recomendado - primera celda de la segunda fila vacía

municipio_nombre	solicitudes_anio	solicitudes_categoria	solicitudes_numero
Municipalidad X	2015	Poda y arbolado	
	2015	Recolección de residuos	198

Recomendado - primera celda de la segunda fila completa

municipio_nombre	solicitudes_anio	solicitudes_categoria	solicitudes_numero
Municipalidad X	2015	Poda y arbolado	340
Municipalidad X	2015	Recolección de residuos	198

Figura 4.4 – Recomendaciones sobre agrupación de conceptos.

En la Figura 4.5, se observan 4 ejemplos de datasets extraídos del portal de datos abiertos de:

- República Argentina [AUD21] llamado "expo-2020-vegetales-certificados-senasa.csv".
- Ciudad de Buenos Aires [BAC21], "centros_de_salud_nivel_1_BADATA_WGS84.csv".
- Ministerio de Justicia y Derechos Humanos de la Argentina [MIN21] llamado "cv-magistrados-poder-judicial-ministerio-publico-20210112.csv".
- Ciudad de Buenos Aires [BAC21] llamado "dataset_flujo_vehicular.csv".

En esta Figura 4.5, se muestran 4 partes, A, B, C y D correspondientemente:

- En la Parte A, en el dataset de ejemplo, se observa que el registro 26394, presenta casos *Nulos*, es decir, sin datos ni espacios en los campos, como se muestra en un círculo rojo.
- En la Parte B, en el dataset de ejemplo, se observa que para los números de registros 8 y 9, presentan casos del mismo tipo que la Parte A, es decir, *Nulos*. Esto se muestra con dos círculos en color rojo.
- En la Parte C, es un ejemplo de dataset en el que, en dicha porción de registros, todos se ven afectados, es decir, existen varios casos del tipo *Nulos*. Esto se muestra en círculos rojos. Este ejemplo denota que pueden existir datasets más afectados que otros.
- En la Parte D, en el dataset de ejemplo, se muestra un caso para el tipo *No Disponibles* (con datos que indiquen "N/D", "N/A", "NULL", "-", "- -" o bien "---" en el conjunto de datos). Se observa que, en todos los registros visualizados, se presentan leyendas "N/D". esto se muestra en círculos rojos.

Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales



expo-2020-vegetales-certificados-senasa.csv

26384	2020-06-08	SAN PEDRO	Buenos Aires	06	ESPAÑA	410	ESP	Europa	Limón Fresco	BODEGA	74.844
26385	2020-06-08	SAN PEDRO	Buenos Aires	06	ESPAÑA	410	ESP	Europa	Limón Fresco	CONT	176.256
26386	2020-06-08	SAN PEDRO	Buenos Aires	06	ESPAÑA	410	ESP	Europa	Grano de Garbanzo	CONT	98.000
26387	2020-06-08	SAN PEDRO	Buenos Aires	06	POLONIA	424	POL	Europa	Grano de Maíz Pisingallo	CONT	48.000
26388	2020-06-08	SAN PEDRO	Buenos Aires	06	ITALIA	417	ITA	Europa	Limón Fresco	CONT	97.200
26389	2020-06-08	SAN PEDRO	Buenos Aires	06	COLOMBIA	205	COL	América	Grano de Maíz Pisingallo	CONT	260.000
26390	2020-06-08	SAN PEDRO	Buenos Aires	06	ECUADOR	210	ECU	América	Grano de Maíz Pisingallo	CONT	230.390
26391	2020-06-08	SAN PEDRO	Buenos Aires	06	ISRAEL	319	ISR	Asia	Grano de Girasol	CONT	21.500
26392	2020-06-08	SAN PEDRO	Buenos Aires	06	CANADA	204	CAN	América	Limón Fresco	CONT	68.040
26393	2020-06-08	SAN PEDRO	Buenos Aires	06	REPÚBLICA DOMINICANA	209	DOM	América	Grano de Maíz Pisingallo	CONT	102.000
26394	2020-06-08	SAN PEDRO	Buenos Aires	06	KOCHABAMBA	262	BOL	América	Limón Fresco	CONT	51.840
26395	2020-06-08	SAN PEDRO	Buenos Aires	06	GHANA	117	GH	África	Grano de Maíz Pisingallo	CONT	128.900
26396	2020-06-08	SAN PEDRO	Buenos Aires	06	SUECIA	113	SWE	Europa	Limón Fresco	CONT	34.125
26397	2020-06-08	SAN PEDRO	Buenos Aires	06	EGIPTO	113	EGY	África	Grano de Maíz Pisingallo	CONT	52.000



centros-de-salud_nivel_1_BADATA_WGS84.csv

8	"POINT (-58.4022498496126 -34.6253913497638)", "38", Centro de Salud Nivel 1 San Cristóbal (CeSAC N° 45), 4941-0969/ 0826/1987, DRA. Natalia Zampichiat, RAMOS MEJIA, I, "Enfermería, Farmacéutico, Fonoaudiología, Ginecología, Kinesología, Medicina general y/o familiar, Nutrición, Pediatría, Psicología, Psiquiatría, Trabajo social", COCHABAMBA, "2622", COCHABAMBA 2622, Más información en: http://www.buenosaires.gov.ar/salud/centro-de-salud-san-cristobal,CeSAC N° 45, Centro de Salud Nivel 1 San Cristóbal, (CeSAC N° 45), CENTRO DE SALUD NIVEL 1, NO, SI, SI, SAN CRISTOBAL, 3.000000000000
9	"POINT (-58.3659808000432 -34.6401070890471)", "37", Centro de Salud Nivel 1 La Boca 1 (CeSAC N° 9), 4302-9983, DRA. BORISIUK LUCÍA, ARGERICH, I, "Clínica médica, Enfermería, Farmacéutico, Fonoaudiología, Ginecología, Kinesología, Medicina general y/o familiar, Medicina general, Nutrición, Obstetricia, Odontología, Pediatría, Psicología, Psicopedagogía, Psiquiatría, Sociología, Terapia ocupacional", IRALA, "1254", IRALA 1254, Más información en: http://www.buenosaires.gov.ar/salud/centro-de-salud-la-boca-1,CeSAC N° 9, Centro de Salud Nivel 1 La Boca 1, (CeSAC N° 9), CENTRO DE SALUD NIVEL 1, SI, SI, SI, SI, BOCA, 4.000000000000
10	"POINT (-58.3973625693378 -34.6560438464025)", "36", Centro de Salud Nivel 1 Barracas 2 (CeSAC N° 8), 4302-3825, DRA. BEATRIZ BELLONI, PENNA, I, "Clínica médica, Farmacéutico, Ginecología, Medicina general y/o familiar, Obstetricia, Pediatría, Psicología, Trabajo social, Nutrición y Psiquiatría", CRUZ OSVALDO, "0", LUNA, "CRUZ, OSVALDO y LUNA", Más información en:



cv-magistrados-poder-judicial-ministerio-publico-20210112.csv

6	"ABDALA, Martín Eugenio", "21328980", "EX-2018-11563790-APN-DDMIP#MJ", "357", "Consejo de la Magistratura de la Nación", "argentina", "Universidad Nacional de Tucumán", "Facultad de Derecho y Ciencias Sociales", "1992-1-24", "2015-09-10", "3285945-ABDALA, MARTÍN EUGENIO", "http://datos.jus.gov.ar/cv/3285945-ABDALA,%20MARTÍN%20EUGENIO.txt", "http://datos.jus.gov.ar/cv/3285945-ABDALA,%20MARTÍN%20EUGENIO.html"
7	"ABRALDES, Sandro Fabio", "2185063", "XP-504:0061242/2015", "104/MPF", "Ministerio Público Fiscal", "1975-04-10", "argentina", "Universidad de Belgrano", "Facultad de Derecho y Ciencias Sociales", "1994-01-24", "2014-06-26", "3231917-Abrales", "http://datos.jus.gov.ar/cv/3231917-Abrales.txt", "http://datos.jus.gov.ar/cv/3231917-Abrales.html"
8	"ABRALDES, Sandro Fabio", "2185063", "XP-504:0061242/2015", "104/MPF", "Ministerio Público Fiscal", "1975-04-10", "argentina", "Universidad de Belgrano", "Facultad de Derecho y Ciencias Sociales", "1994-01-24", "2014-06-26", "310778-Abrales", "http://datos.jus.gov.ar/cv/310778-Abrales.txt", "http://datos.jus.gov.ar/cv/310778-Abrales.html"
9	"ABRAMOVICH COSARÍN, Víctor Ernesto", "16554338", "EXP-504:0050359/2014", "09/MPF", "Ministerio Público Fiscal", "1965-01-25", "argentina", "Buenos Aires", "Universidad de Buenos Aires", "Facultad de Derecho y Ciencias Sociales", "1989-07-03", "2722205-Abramovich Cosarín", "http://datos.jus.gov.ar/cv/2722205-Abramovich%20Cosarín.txt", "http://datos.jus.gov.ar/cv/2722205-Abramovich%20Cosarín.html", "06"
10	"ABREUT DE BEGHER, Liliana Edith", "14038375", "XP-504:0060402", "argentina", "Ciudad Autónoma de Buenos Aires", "Ciudad Autónoma de Buenos Aires", "Universidad de Buenos Aires", "Facultad de Derecho y Ciencias Sociales", "1985-10-11", "1095-ABREUT", "http://datos.jus.gov.ar/cv/10095-ABREUT.txt", "http://datos.jus.gov.ar/cv/10095-ABREUT.html"



flujo-vehicular-2020.csv

2527	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	ILLIA	N/D	violación, Violación	Normal	Centro	2.0
2528	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	ILLIA	N/D	violación, Violación	Normal	Provincia	10.0
2529	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	ILP	N/D	violación, Violación	Normal	Centro	23.0
2530	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	ILP	N/D	violación, Violación	Normal	Provincia	23.0
2531	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	PE	N/D	violación, Violación	Normal	Provincia	27.0
2532	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	PB	N/D	violación, Violación	Normal	Centro	26.0
2533	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	PB	N/D	violación, Violación	Normal	Provincia	21.0
2534	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	PB	N/D	violación, Violación	Normal	Centro	23.0
2535	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	SALGUERO	N/D	violación, Violación	Normal	Provincia	13.0
2536	2020-01-01	00:00:00	enero	Miércoles	1.0	0.0	1.0	SARMIENTO	N/D	violación, Violación	Normal	Provincia	10.0
2537	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	ALBERDI	N/D	violación, Violación	Normal	Provincia	1.0
2538	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	AVELLA NEDA	N/D	violación, Violación	Normal	Centro	6.0
2539	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	AVELLA NEDA	N/D	violación, Violación	Normal	Provincia	6.0
2540	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	DELLEPIANE LINIERS	N/D	violación, Violación	Normal	Centro	1.0
2541	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	DELLEPIANE LINIERS	N/D	violación, Violación	Normal	Provincia	2.0
2542	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	ILLIA	N/D	violación, Violación	Normal	Provincia	8.0
2543	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	ILP	N/D	violación, Violación	Normal	Centro	33.0
2544	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	ILP	N/D	violación, Violación	Normal	Provincia	8.0
2545	2020-01-01	00:00:00	enero	Miércoles	1.0	1.0	2.0	PB	N/D	violación, Violación	Normal	Provincia	28.0



Figura 4.5 – Datasets para la métrica de datos faltantes y/o completos.



4.1.4. Caracteres Inválidos

Esta métrica propuesta permite identificar los caracteres especiales del conjunto de datos analizado. Se podría incluir el carácter afectado, y el número de registro del dataset, como así también, el nombre de la columna/campo en el que aparece.

Es importante localizar este tipo de caracteres con el fin de no alterar la identificación y análisis de los valores contenidos en los conjuntos de datos. El gran problema que conlleva que los datos se presenten con caracteres inválidos, supondrá una pérdida de información y, por consiguiente, una pérdida de la objetividad de lo que se está analizando como resultado.

Al igual que la métrica de detección de datos faltantes y/o completos de la sección anterior, para el análisis de esta métrica, se toman en cuenta medidas cuantitativas de calidad de datos.

Para este aspecto, la dimensión de calidad de datos utilizada es: Exactitud. Es decir, se verifica si los datos son exactos/correctos en base a un aspecto puntual que se relaciona con el factor, Correctitud sintáctica, esto es, si los datos se encuentran libres de errores sintácticos. Esto puede darse por un carácter inválido/especial.

En la Figura 4.6, se muestra un ejemplo del dataset extraído del portal de datos abiertos del Ministerio de Agricultura, Ganadería y Pesca [AGP21], llamado “*produccion-de-bioetanol-por-insumo-.csv*”.

```
1 32,Argentina,2017,Enero,m3,metros cubicos,37794,46551,84345
2 32,Argentina,2017,Febrero,m3,metros cubicos,35607,44489,80096
3 32,Argentina,2017,Marzo,m3,metros cubicos,38696,44100,82796
4 32,Argentina,2017,Abril,m3,metros cubicos,35355,45218,80573
5 32,Argentina,2017,Mayo,m3,metros cubicos,44053,40222,84275
6 32,Argentina,2017,Junio,m3,metros cubicos,51159,41800,92959
7 32,Argentina,2017,Julio,m3,metros cubicos,53924,43396,97320
8 32,Argentina,2017,Agosto,m3,metros cubicos,55711,48852,104563
9 32,Argentina,2017,Septiembre,m3,metros cubicos,55345,47483,102828
10 32,Argentina,2017,Octubre,m3,metros cubicos,60415,50850,111265
11 32,Argentina,2017,Noviembre,m3,metros cubicos,48724,47841,96565
12 32,Argentina,2017,Diciembre,m3,metros cubicos,36361,51160,87521
13 32,Argentina,2018,Enero,m3,metros cubicos,33251,50443,83694
14 32,Argentina,2018,Febrero,m3,metros cubicos,35328,45593,80921
15 32,Argentina,2018,Marzo,m3,metros cubicos,17968,49519,67487
16 32,Argentina,2018,Abril,m3,metros cubicos,24522,50537,75059
17 32,Argentina,2018,Mayo,m3,metros cubicos,38336,49606,87942
18 32,Argentina,2018,Junio,m3,metros cubicos,57254,49484,106738
19 32,Argentina,2018,Julio,m3,metros cubicos,60456,46815,107271
20 32,Argentina,2018,Agosto,m3,metros cubicos,59362,47953,107315
21 32,Argentina,2018,Septiembre,m3,metros cubicos,59862,46146,106008
22 32,Argentina,2018,Octubre,m3,metros cubicos,54597,48608,103205
23 32,Argentina,2018,Noviembre,m3,metros cubicos,49554,48981,98535
24 32,Argentina,2018,Diciembre,m3,metros cubicos,37673,51933,89606
25 32,Argentina,2019,Enero,m3,metros cubicos,30505,39815,70320
26 32,Argentina,2019,Febrero,m3,metros cubicos,36711,41695,69396
```

Figura 4.6 – Dataset para la Métrica de caracteres especiales.



Para los casos de caracteres especiales de la Figura 4.6, se han detectado los problemas de tipeo existentes distinguido por columnas/campos del dataset:

- Para la columna “*nom_unimed*” existen 2 casos.
- Para la columna “*mes*” existe un caso.
- Para la columna “*cod_unimed*” existe un caso.

Siendo:

- Registro Número 3, contiene el carácter “~” en la columna “*cod_unimed*”.
- Registro Número 9, contiene el carácter “μ” en la columna “*nom_unimed*”.
- Registro Número 23, contiene el carácter “~” en la columna “*mes*”.
- Registro Número 26, contiene el carácter “~” en la columna “*nom_unimed*”.

4.2. Métricas No Críticas

A continuación, se presentan cada una de las métricas propuestas en esta categoría.

4.2.1. Redundancia para el dominio de una columna

Esta métrica propuesta permite calcular la cantidad de valores repetidos en una misma columna/campo, es decir, trabaja con el dominio de datos (conjunto de valores posibles).

Para analizar este aspecto, se muestra la cantidad de columnas afectadas que poseen valores iguales entre si sobre la cantidad total de registros del conjunto de datos:

Dentro de los estándares de calidad, recomendado por el sitio gubernamental de datos abiertos de la República Argentina [AUD21], se sugiere que las entidades que aparezcan entre los datos de un campo textual deben tener una descripción única. Por ello, la importancia de detectar casos de valores igual, con el fin de saber si están bien agregados o deben ser modificados para que cumplan con una misma descripción. Por lo que se sugiere que toda mención que se realice a una entidad dada debe hacerse usando exactamente la misma cadena de caracteres cada vez [TIP21]:

- Las **descripciones de entidades** deberían elegirse siempre de forma tal que cumplan con el estándar específico que las describe, en caso de que este exista.
- **Cuando este estándar no existe** y hay dudas respecto del criterio a adoptar para elegir la descripción única de una entidad, debe **privilegiarse siempre aquella que sea lo más explícita**, descriptiva y declarativa posible.

Un ejemplo de esta descripción de aspecto es la que se muestra en la Figura 4.7 extraída de [TIP21], allí se puede observar que los cuatro valores de texto refieren a la misma entidad, siendo:

- Ciudad Autónoma de Buenos Aires
- CABA



- Capital Federal
- Ciudad de Buenos Aires

En la parte inferior de la Figura 4.7 extraída de [TIP21], puede observarse que la descripción estándar recomendada por la guía de buenas prácticas [GUI21], es “Ciudad Autónoma de Buenos Aires”. Este presenta un agrupamiento lógico definido, que permitirá un mejor análisis a futuro sobre la información establecida. Es decir, siempre que sea posible, es importante incorporar un estándar para identificar un marco común en el tratamiento de un aspecto dado. Para más información, ver la *Guía para la identificación y uso de entidades interoperables* [GUI21].

Al igual que en la sección 4.1.2., esta métrica se relaciona con la dimensión de Unicidad. En dicha sección, se explica esta dimensión. Para esta dimensión de calidad, también se utiliza el factor de calidad: No-duplicación en el cual se analiza el grado de repetición de valores de un mismo campo.

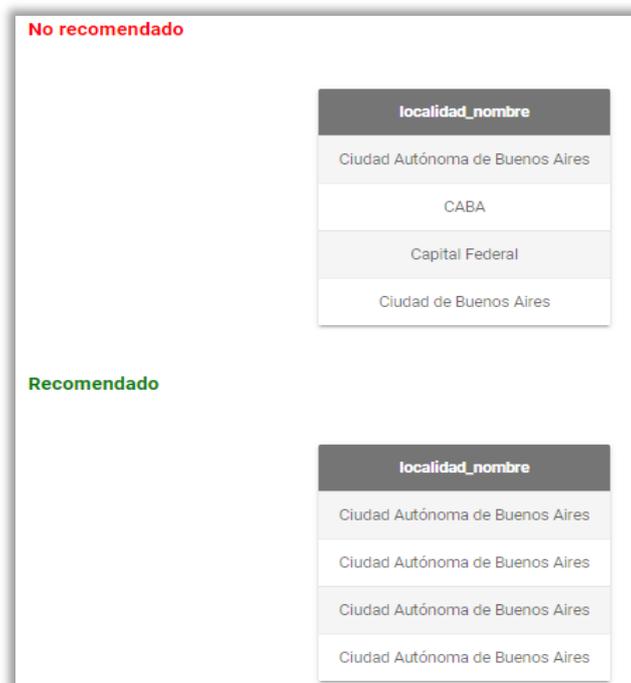


Figura 4.7 – Buenas prácticas de las descripciones de datos abiertos.

En la Figura 4.8, se muestra un ejemplo de un dataset que fue extraído del portal de datos abiertos de la Ciudad de Buenos Aires [BAC21], llamado “padron-de-operadores-organicos-certificados.csv”, en el que se lleva a cabo un análisis de casos de redundancia en el dominio de valores de una misma columna. Para ello se visualiza la cantidad de columnas detectadas que tienen repetición en sus datos, sobre el total de columnas del dataset.

Por ejemplo:

- Se puede observar que el valor “32”, se repite en varios registros para la misma columna/campo, “pais_id”. Esto se muestra en círculo en color azul.
- Se puede observar que el valor “ARGENTINA”, se repite en varios registros para la misma columna/campo, “pais”. Esto se muestra en círculo en color amarillo.



- Se puede observar que el valor “BUENOS AIRES”, se repite en varios registros para la misma columna/campo, “provincia”. Esto se muestra en círculo en color rosa. La misma redundancia se observa en el campo “provincia_id” con el valor 6 (tercer dato).

```
padron-de-operadores-organicos-certificados.csv
1 pais_id,pais,provincia_id,provincia,departamento,localidad,rubro,productos,categori
a_id,categoria_desc,Certificadora_id,certificadora_deno,razón
social,establecimiento
2 32 ARGENTINA 6 BUENOS AIRES BARADERO,INDEFINIDA,AGRICULTURA,SOJA,1,Productores,2,OIA S.A.,ADD CROPS
S.A.,BARADERO 1
3 32 ARGENTINA 6 BUENOS AIRES BARADERO,INDEFINIDA,AGRICULTURA,"GIRASOL, MIJO,
ALPISTE,CETIL",1,Productores,2,OIA S.A.,ADD CROPS S.A.,BARADERO 2
4 32 ARGENTINA 6 BUENOS AIRES BARADERO,INDEFINIDA,AGRICULTURA,"MAIZ, AVENA, CAMPO
NATURA",1,Productores,2,OIA S.A.,ADD CROPS S.A.,EL BUEN RETIRO
5 32 ARGENTINA 6 BUENOS AIRES AZUL,INDEFINIDA,AGRICULTURA,"SOJA,
ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO
6 32 ARGENTINA 6 BUENOS AIRES AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE,
TRIGO",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL HORIZONTE
7 32 ARGENTINA 6 BUENOS AIRES CARMEN DE
PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO
ROSSI S.H. DON GTSLEMO
8 32 ARGENTINA 6 BUENOS AIRES CARMEN DE
PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO
ROSSI S.H. ONELIA
9 32 ARGENTINA 6 BUENOS AIRES BARADERO,INDEFINIDA,FRUTICULTURA/HORTICULTURA,"NUEZ
PECAN, PASTURAS, INCULTO",1,Productores,2,OIA S.A.,ALVAREZ ANA MARIA,EL PAMPA
ORGANICO
10 32 ARGENTINA 6 BUENOS AIRES VILLARINO,INDEFINIDA,GANADERIA,GANADERIA
OVINA,1,Productores,1,Angercert S.R.L.,ANIJG AGENCY S.A.,EL FORTIN
11 32 ARGENTINA 6 BUENOS AIRES AZUL,INDEFINIDA,AGRICULTURA,"SOJA,
ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO
12 32 ARGENTINA 6 BUENOS AIRES MERCEDES,INDEFINIDA,FRUTICULTURA/
HORTICULTURA,ARANDANO,1,Productores,2,OIA S.A.,ARBE JUAN LUCAS ,NATURAL PAMPA
```

Figura 4.8 – Pantalla de Medición de Redundancia en el dominio de valores de una columna.

Esto quiere decir que existen varios casos en el que el dato es el mismo para una determinada columna, por lo que, si en todas sus columnas presenta el mismo valor, podría llegar a omitirse y así ocupar un menor tamaño de archivo para un dataset, evitar tiempos de demora en procesamiento y lectura de análisis.

4.2.2. Redundancia entre campos de una misma fila

Esta métrica propuesta permite identificar la cantidad de casos con redundancia entre los valores de los campos para un mismo registro.

Un principio importante para la calidad de datos es no repetir los mismos valores en más de una columna para la misma fila del dataset. Esto es, debe estar bien definida y descripta cada columna que se brinda en el conjunto de datos, ya que representa un valor específico en el análisis lógico y representativo de los datos en formato abierto.

“Una de las operaciones a las que con mayor frecuencia se enfrenta cualquier desarrollador de bases de datos es la detección y tratamiento de datos duplicados, es decir, encontrar varios registros exactamente iguales en una tabla, debido a problemas en el diseño e inconsistencias de



la propia base de datos, o bien para localizar determinados subconjuntos de datos con condiciones que se repitan dentro de una misma tabla” [SQL15]. Para mejorar la calidad de los datos, es necesario eliminar la información redundante o repetitiva. La duplicación de los datos puede conllevar a equivocaciones o bien errores lógicos en el análisis final que pueden ser consecuencia de no contar con un enfoque integrado en la lógica del dataset.

Como buena práctica, el sitio gubernamental de datos abiertos de la República Argentina [AUD21], sugiere como una de las pautas para tener en cuenta en la construcción de los datasets es que “no se deben repetir nombres entre los campos” [CSV21].

Al igual que en la sección 4.1.2., esta métrica se relaciona con la dimensión de Unicidad. En dicha sección, se explica esta dimensión de calidad. También se utiliza el factor de calidad: No-duplicación en el cual se analiza la cantidad de casos con redundancia entre los valores de los campos para un mismo registro.

En la Figura 4.9, se muestra un ejemplo de un dataset que fue extraído del portal de datos abiertos de la Ciudad de Buenos Aires [BAC21], llamado “padron-de-operadores-organicos-certificados.csv”, para la medición de redundancia entre los campos de una misma fila. Por ejemplo, se observa que el valor “CIUDAD AUTONOMA BUENOS AIRES”, se muestra en más de una columna de una misma fila, para los números de registros desde el 1289 al 1301 inclusive.

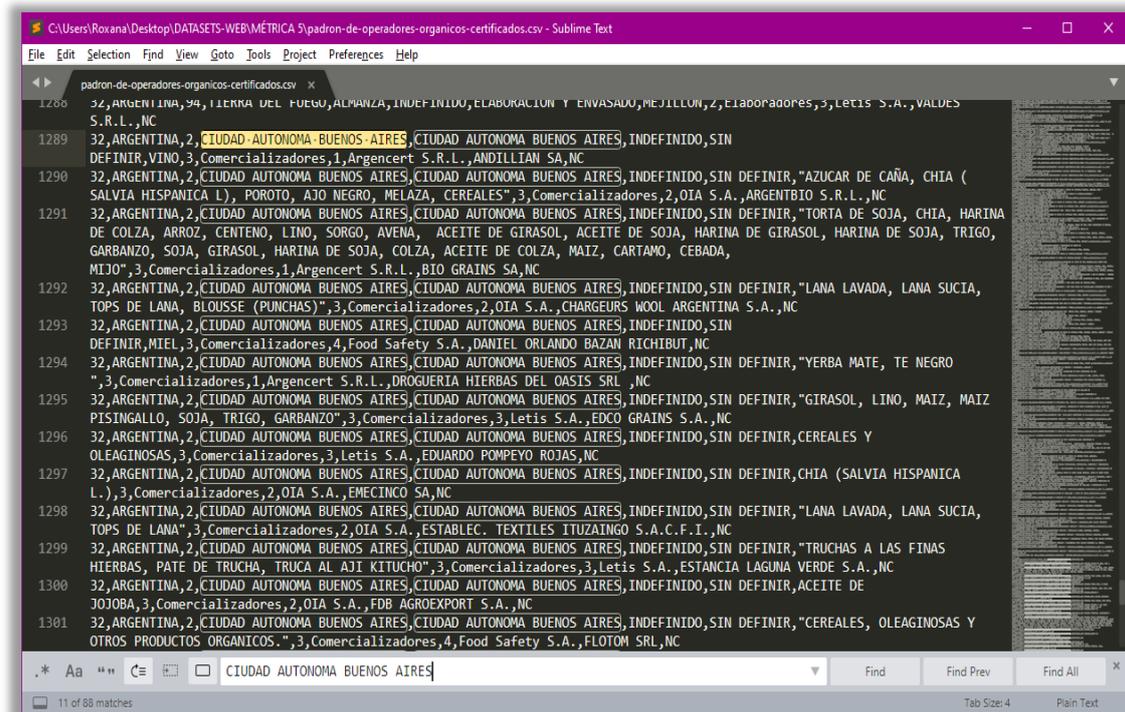


Figura 4.9 – Dataset para Redundancia entre campos de una misma fila (registros 1289 al 1301).

Continuando con el mismo dataset de la Figura 4.9, en la Figura 4.10 se muestra el valor “CORDOBA”. Si bien se repite en una misma columna, existen 5 casos en los que se repite para más de una columna/campo de los números de registros 1346, 1347, 1353, 1356 y 1358 de la figura. Esto se muestra con indicadores en color rojo en los números de registros.



```
padron-de-operadores-organicos-certificados.csv x
1345 32,ARGENTINA,10,CATAMARCA,SAN FERNANDO DEL VALLE DE CATAMARCA,INDEFINIDO,SIN DEFINIR,ACEITE DE
OLIVA,3,Comercializadores,1,Argencert S.R.L.,REGINA SA,NC
1346 32,ARGENTINA,14,CORDOBA,CORDOBA,INDEFINIDO,SIN DEFINIR,COMERCIALIZACION DE CEREALES Y
OLEAGINOSAS,3,Comercializadores,4,Food Safety S.A.,CONOSUR FOODS ARGENTINA S.A.,NC
1347 32,ARGENTINA,14,CORDOBA,CORDOBA,INDEFINIDO,SIN DEFINIR,MIEL Y SUBPRODUCTOS DE LA
COLMENA,3,Comercializadores,4,Food Safety S.A.,FRANCO BATTAN NERY,NC
1348 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,"MANI, MANI CONFITERIA, MANI FRITO SALADO, MANI TOSTADO,
ACEITE DE MANI, PASTA DE MANI, HARINA DE MANI",3,Comercializadores,2,OIA S.A.,JL S.A.,NC
1349 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,MANI,3,Comercializadores,4,Food Safety S.A.,JL SA,NC
1350 32,ARGENTINA,14,CORDOBA,BARRIO VILLA BELGRANO,INDEFINIDO,SIN DEFINIR,"ESPARRAGO , MIEL, VINO,
DULCE",3,Comercializadores,1,Argencert S.R.L.,LA MECO S.R.L,NC
1351 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,"MANI, MANI CONFITERIA, MANI FRITO SALADO, MANI TOSTADO,
ACEITE DE MANI, HARINA DE MANI",3,Comercializadores,2,OIA S.A.,LORENZATI RUETSCH Y CIA. S.A.,NC
1352 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,MANI,3,Comercializadores,4,Food Safety S.A.,"LORENZATI,
RUETSCH Y CIA - LR&C SA",NC
1353 32,ARGENTINA,14,CORDOBA,CORDOBA,INDEFINIDO,SIN DEFINIR,COMERCIALIZACION DE MANI Y SUS
DERIVADOS,3,Comercializadores,4,Food Safety S.A.,LORENZO PERLO Y CIA SA,NC
1354 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,"MANI, MANI CONFITERIA, MANI FRITO SALADO, MANI TOSTADO,
ACEITE DE MANI, HARINA DE MANI, PASTA DE MANI",3,Comercializadores,2,OIA S.A.,NUTRIN S.A.,NC
1355 32,ARGENTINA,14,CORDOBA,TICINO,INDEFINIDO,SIN DEFINIR,MANI,3,Comercializadores,4,Food Safety S.A.,NUTRIN SA,NC
1356 32,ARGENTINA,14,CORDOBA,CORDOBA CAPITAL,INDEFINIDO,SIN DEFINIR,MANI,3,Comercializadores,2,OIA S.A.,PRONUT
S.R.L.,NC
1357 32,ARGENTINA,14,CORDOBA,GENERAL SAN MARTIN,INDEFINIDO,SIN DEFINIR,MANI Y SOJA,3,Comercializadores,4,Food
Safety S.A.,PRONUT SRL,NC
1358 32,ARGENTINA,14,CORDOBA,CORDOBA,INDEFINIDO,SIN DEFINIR,CEREALES Y OLEAGINOSAS,3,Comercializadores,4,Food
Safety S.A.,BAFAGOS SA,NC
```

Figura 4.10 – Dataset para Redundancia entre campos de una misma fila (registros 1346 al 1358).

4.2.3. Detección de valores ID

Esta métrica permite realizar una estimación sobre la cantidad de campos “ID” identificados en las columnas del dataset. Para ello, se analizan los nombres que contengan, tanto en mayúsculas como en minúsculas, los siguientes elementos:

- 'id',
- 'id_'
- '_id'

Los campos que contienen en sus descripciones “ID”, son utilizados con valores numéricos y representan un código o valor único de enteros que no es nulo, y, además, se implementan para identificar de forma unívoca a cada una de las filas del conjunto de datos. Resulta necesario cuantificar la cantidad de campos para este tipo, ya que, si bien son identificadores, a los ciudadanos y/u organismos que acceder a este conjunto de datos, podrían no interpretar el significado del código numérico que se muestra, esto es, en muchos casos los datasets forman parte de un análisis de diversos estudios estadísticos, que no logran detectar el significado y/o utilidad de la nomenclatura tradicional de códigos ID. Esto se resolvería, si en el sitio oficial del cual fue extraído el dataset, se acompaña el diccionario de datos correspondiente, con el fin de comprender los campos del conjunto de datos, y, sobre todo, el significado de estos campos ID.

Si bien los datasets abiertos, deben contener datos bien definidos, organizados y justificados, como buena práctica, el sitio gubernamental de datos abiertos de la República Argentina [AUD21], sugiere la utilización de un campo identificador en el dataset, debido a que “suele ser útil para la



identificación unívoca de variables en algunos sistemas o aplicaciones, pero no en la generalidad de los casos” [PAN21]. Cabe aclarar que no todos los conjuntos de datos tienen campos ID, esto es de forma opcional.

En algunos datasets de Argentina, el ID es el identificador único del catálogo dentro de la Red de Nodos de Datos Abiertos de la Administración Pública Nacional. Este identificador es otorgado por la Dirección Nacional de Datos e Información Pública cuando un nuevo nodo pide ser incorporado a la red para su federación en el nodo concentrador de datos abiertos de la Administración Pública Nacional, APN [IDA17], como buena práctica se sugiere que sea una o más palabras en minúsculas, separadas con guiones medios, sin usar caracteres especiales.

Desde el portal de datos abiertos de la República Argentina, se presenta una guía orientativa, que busca ayudar a los organismos a instrumentar la Política de Datos Abiertos impulsada desde el Gobierno de la Nación Argentina, a través del Decreto N° 117/2016 del 12 de enero de 2016. “Esta es una guía de buenas prácticas para el uso de entidades interoperables. Se trata de datos básicos y fundamentales cuyo uso se repite frecuentemente entre datasets de temáticas y fuentes distintas. Para hacer estas recomendaciones, nos basamos en estándares usados a nivel nacional e internacional y en la experiencia de trabajo del equipo de la Dirección Nacional de Datos e Información Pública del Secretaría de Gobierno de Modernización de la Jefatura de Gabinete de Ministros de la Nación. Esta es una guía colaborativa y en progreso. Valoramos, y alentamos, a organizaciones y ciudadanos a plantear ideas, sugerencias, y comentarios que nos ayuden a crear un mejor documento” [GUI21].

Es importante destacar que los datasets incluyen campos de ID que son los que permiten que el conjunto de datos sea interoperable entre varios datasets. “Las entidades interoperables son las que permiten que los datasets hablen entre sí, pero esto no puede suceder cuando dos datasets nombran de forma distinta a una misma entidad interoperable (como cuando se usan distintos sistemas de id o se nombra una misma entidad con/sin mayúsculas, usando artículos y preposiciones (o no usándolos), usando abreviaturas, siglas, tildes, forma corta o completa de un nombre, etc. Para que los datasets puedan ser interoperables, deben identificarse todas las entidades interoperables presentes en un dataset y asegurarse de que los datos sobre ellas siguen el mismo estándar” [EST21]. A continuación, se muestran los tipos de entidades interoperables, cabe destacar que el tratamiento al detalle de las entidades interoperables no es parte del alcance de esta tesis. Este tema es abordado en [GUI21].

En la Figura 4.12, se muestra un ejemplo de un dataset que fue extraído del portal de datos abiertos de la Ciudad de Buenos Aires [BAC21], llamado “*padron-de-operadores-organicos-certificados.csv*”, para la estimación de campos con ID, para la detección de las columnas que contienen 'id', 'id_' y/o '_id', tanto en mayúsculas como en minúsculas. Allí se observan los Títulos/Campos detectados:

- pais_id
- id
- provincia_id
- categoria_id
- Certificadora_id



```
padron-de-operadores-organicos-certificados - copia.csv x
1  pais_id,id,provincia_id,provincia,departamento,locc,rubro,productos,categoria_id,
   categoria_desc,Certificadora_id,certificadora_deno,razón social,establecimiento
2  32,ARGENTINA,6,BUENOS
   AIRES,BARADERO,INDEFINIDA,AGRICULTURA,SOJA,1,Productores,2,OIA S.A.,ADD CROPS
   S.A.,BARADERO 1
3  32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,AGRICULTURA,"GIRASOL, MIJO,
   RESIDUO VEGETAL",1,Productores,2,OIA S.A.,ADD CROPS S.A.,BARADERO 2
4  32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,AGRICULTURA,"MAIZ, AVENA, CAMPO
   NATURAL",1,Productores,2,OIA S.A.,ADD CROPS S.A.,EL BUEN RETIRO
5  32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA,
   ALPISTE",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL RELINCHO
6  32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA, ALPISTE,
   TRIGO",1,Productores,3,Letis S.A.,AGRICOLA EL HORIZONTE S.A.,EL HORIZONTE
7  32,ARGENTINA,6,BUENOS AIRES,CARMEN DE
   PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO
   ROSSI S.H.,DON GISLENO
8  32,ARGENTINA,6,BUENOS AIRES,CARMEN DE
   PATAGONES,INDEFINIDA,AGRICULTURA,CEREALES,1,Productores,4,Food Safety S.A.,AGRO
   ROSSI S.H.,ONELIA
9  32,ARGENTINA,6,BUENOS AIRES,BARADERO,INDEFINIDA,FRUTICULTURA/HORTICULTURA,"NUEZ
   PECAN, PASTURAS, INCULTO",1,Productores,2,OIA S.A.,ALVAREZ ANA MARIA,EL PAMPA
   ORGANICO
10 32,ARGENTINA,6,BUENOS AIRES,VILLARINO,INDEFINIDA,GANADERIA,GANADERIA
   BOVINA,1,Productores,1,Argencert S.R.L.,ANIJG AGENCY S.A.,EL FORTIN
11 32,ARGENTINA,6,BUENOS AIRES,AZUL,INDEFINIDA,AGRICULTURA,"SOJA,
```

Figura 4.11 – Dataset para la detección de casos del tipo ID.

Éstos representan los códigos identificadores que permitirán la interoperabilidad entre datasets, por lo que es importante detectarlos para lograr conocer el propósito de cada uno de éstos y así facilitar a un correcto análisis de los datos abiertos.

4.2.4. Campos Triviales

Esta métrica propuesta permite identificar y mostrar si existen campos con valores iguales para una misma columna. En caso de existir, se indica la cantidad de columnas afectadas sobre el total de columnas del dataset y los correspondientes campos y sus valores encontrados.

Un registro duplicado ocurre cuando el mismo dato se ha introducido más de una vez, por lo que es importante, detectar si existen campos/columnas que poseen en todos sus valores el mismo dato. El descubrimiento de estos casos permitirá conocer si hay campos que pueden ser omitidos en el dataset, ya que éstos podrían indicarse como dato en el nombre del conjunto de datos.

Por ejemplo: si en todos los registros se detectara un campo país = Argentina, entonces el dataset debería contener en su nombre Argentina, siendo:

- Dataset llamado “Casos registrados de Covid-19”, podría llamarse “Casos registrados de Covid-19 en Argentina”



Al igual que en la sección 4.1.2., esta métrica se relaciona con la dimensión de Unicidad. En dicha sección, se explica esta dimensión. Para esta dimensión de calidad, se utiliza el factor de calidad: duplicación en el cual se analiza el grado de repetición de valores de un mismo campo.

En la Figura 4.13, se muestra un ejemplo de un dataset que fue extraído del portal de datos abiertos del Ministerio de Salud de Argentina [SAL21], llamado “*dengue-zika-nacional-201807.csv*”, para el análisis de campos triviales, en el que se observa el campo “anio”, con el mismo valor en todas sus filas, es decir “2018”. Esto se muestra con un círculo en color rojo. Debido a que el valor siempre es el mismo, se podría omitir e incluir dicho valor al nombre del dataset, por ejemplo: en lugar de tener un dataset “*dengue-zika-nacional-201807.csv*”, se podría llamar “*dengue-zika-nacional-año2018.csv*”.

```
1 "departamento_id", "departamento_nombre", "provincia_id", "provincia_nombre", "anio", "semanas_epidemiologicas", "evento_nombre", "grupo_edad_id", "grupo_edad_desc", "cantidad_casos"
2 999, "*sin dato*", 6, "Buenos Aires", 2018, 11, "Dengue", 10, "De 45 a 64 años", 1
3 35, "Avellaneda", 6, "Buenos Aires", 2018, 7, "Dengue", 8, "De 20 a 24 años", 1
4 35, "Avellaneda", 6, "Buenos Aires", 2018, 7, "Dengue", 10, "De 45 a 64 años", 1
5 35, "Avellaneda", 6, "Buenos Aires", 2018, 9, "Dengue", 6, "De 10 a 14 años", 1
6 35, "Avellaneda", 6, "Buenos Aires", 2018, 9, "Dengue", 8, "De 20 a 24 años", 1
7 35, "Avellaneda", 6, "Buenos Aires", 2018, 9, "Dengue", 9, "De 25 a 34 años", 1
8 35, "Avellaneda", 6, "Buenos Aires", 2018, 10, "Dengue", 8, "De 20 a 24 años", 1
9 35, "Avellaneda", 6, "Buenos Aires", 2018, 10, "Dengue", 10, "De 45 a 64 años", 1
10 35, "Avellaneda", 6, "Buenos Aires", 2018, 11, "Dengue", 6, "De 10 a 14 años", 2
11 35, "Avellaneda", 6, "Buenos Aires", 2018, 11, "Dengue", 8, "De 20 a 24 años", 1
12 35, "Avellaneda", 6, "Buenos Aires", 2018, 11, "Dengue", 9, "De 25 a 34 años", 1
13 35, "Avellaneda", 6, "Buenos Aires", 2018, 11, "Dengue", 10, "De 45 a 64 años", 1
14 35, "Avellaneda", 6, "Buenos Aires", 2018, 12, "Dengue", 6, "De 10 a 14 años", 1
15 35, "Avellaneda", 6, "Buenos Aires", 2018, 12, "Dengue", 10, "De 35 a 44 años", 2
16 35, "Avellaneda", 6, "Buenos Aires", 2018, 13, "Dengue", 9, "De 25 a 34 años", 2
17 35, "Avellaneda", 6, "Buenos Aires", 2018, 13, "Dengue", 10, "De 35 a 44 años", 2
18 35, "Avellaneda", 6, "Buenos Aires", 2018, 13, "Dengue", 11, "Mayor o igual de 65 años", 1
19 35, "Avellaneda", 6, "Buenos Aires", 2018, 14, "Dengue", 5, "De 5 a 9 años", 1
20 35, "Avellaneda", 6, "Buenos Aires", 2018, 14, "Dengue", 6, "De 10 a 14 años", 2
21 35, "Avellaneda", 6, "Buenos Aires", 2018, 14, "Dengue", 8, "De 20 a 24 años", 1
22 35, "Avellaneda", 6, "Buenos Aires", 2018, 14, "Dengue", 9, "De 25 a 34 años", 3
```

Figura 4.12 – Dataset para el análisis de campos triviales.

A continuación, se presenta el Capítulo 5, en el que se desarrollan los aspectos de la validación empírica junto con los resultados arrojados para la propuesta general de la presente tesis.



Capítulo 5 – Implementación

En el capítulo anterior se presentaron las métricas de calidad propuestas para la validación de los datos abiertos. En este capítulo, se explicarán como dichas métricas son implementadas en una herramienta desarrollada, la cual es nombrada “HEVDA, *herramienta de validación de datos abiertos*”. Mediante esta aplicación, se realizarán diversas pruebas a datasets seleccionados de algunos portales gubernamentales de datos abiertos más relevantes de Argentina, los cuales formarán parte de la recolección de datos definidos como muestra. Finalmente, se incluye un análisis detallado de los resultados extraídos.

5.1. Herramienta Desarrollada - HEVDA

Como parte de la propuesta de este trabajo, se presenta una herramienta desarrollada que permite la validación de las distintas métricas sugeridas para un conjunto de datos abiertos en formatos CSV (valores separados por comas). 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 práctica orientativa para la corrección de este. Esta herramienta de desarrollo propio fue parte del trabajo de investigación elaborado y publicado en una revista internacional [MRO21].

5.1.1. Aspectos Técnicos

La herramienta de validación fue realizada con el entorno de desarrollo integrado (IDE), Visual Studio Community 2019 [COM19], que es una herramienta completa para programar, depurar, probar e implementar soluciones en cualquier plataforma. Otro de los motivos por el cual fue seleccionada, es que posee un entorno de programación amigable, y, además, su versión Community es gratuita. Por otra parte, existen foros a nivel soporte técnico de la plataforma y material de respaldo.

En cuanto al lenguaje de programación se utilizó C# que es el lenguaje de programación orientado a objetos, con ASP .NET que es el marco proporcionado por Microsoft para el desarrollo web. Puntualmente se utilizó C# junto con ASP .NET para la codificación del lado del servidor, y HTML, CSS, JS para el lado del cliente. Como servidor web, se utilizó Internet Information Services (IIS) versión 10.0.19041.1.



5.1.2. Aspectos Funcionales

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 la elección del tipo de formato a considerar se realizó un estudio producto del cual se publicó un artículo [MAR20], en el que se tomó como caso de muestra el portal gubernamental Argentina Unida [AUD21] con sus 973 datasets a julio del año 2020. Sus resultados concluyeron que el formato más utilizado es el tipo CSV con un 61,6% de uso, de allí dicha elección.

Las funcionalidades detalladas de la herramienta programada son:

- Identificación del cumplimiento de las métricas críticas;
- Detección y detalle de los casos que no cumplen con el formato válido para el tipo de dato decimal;
- Cálculo estimativo de los tipos de datos de los campos del dataset validado;
- Cálculo de la cantidad y porcentaje de los registros duplicados;
- Detalle de los registros duplicados;
- Cálculo de la cantidad y porcentaje de los registros completos;
- Cálculo de la cantidad de casos que poseen campos con registros Nulos (Sin Datos ni espacios en los campos);
- Cálculo de la cantidad de casos que poseen campos con registros Vacíos (Sin Datos y con espacios en los campos);
- Cálculo de la cantidad de casos que poseen campos con registros No Disponibles (Con datos que indique N/D, N/A, NULL, -, - -, --);
- Visualización del detalle de los casos con registros Nulos, Vacíos y No Disponibles;
- Cálculo de la cantidad de columnas afectadas con caracteres especiales y su correspondiente detalle;
- Cálculo de la cantidad de columnas afectadas con valores repetidos en un mismo campo (dominio de valores);
- Detalle con filtros de búsqueda para los campos del dataset y palabras de los campos detectados/as en los casos que hubo registros con valores repetidos para un mismo campo (dominio de valores);
- Cálculo de la cantidad y porcentaje de casos detectados con redundancia entre los valores de los campos para un mismo registro;
- Filtro de búsqueda para los casos detectados con datos redundantes entre los valores de los campos para un mismo registro y su correspondiente detalle;
- Estimación de la cantidad de ID identificados en las columnas del dataset, y la visualización de éstos;
- Cálculo de la cantidad de columnas afectadas con posibles campos triviales y su identificación.



Para más información sobre funcionalidades secundarias de la herramienta HEVDA, se acompaña el [Anexo C: “Funciones secundarias de la Herramienta HEVDA”](#).

5.1.3. Aspectos Bloqueantes

Al efectuar un análisis en un conjunto de datos se pueden encontrar una serie de aspectos que impiden el correcto análisis de validaciones. Estos aspectos son de carácter bloqueante frente a una validación específica, por lo que existen algunos puntos para tener en cuenta que pueden ser de carácter bloqueante o inconvenientes a identificar que pueden complicar a una correcta validación con el dataset.

Previo al comienzo del análisis con la herramienta HEVDA, un archivo dataset no deben cumplir con una serie de condiciones a las que se denominan “bloqueantes”, esto es, en caso de que exista algún tipo de bloqueante en la estructura del archivo, éste no podrá ser validado por el prototipo, ya que impedirá el procesamiento y análisis de las métricas propuestas del conjunto de datos.

A continuación, se identifican los tipos de bloqueantes:

- **EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA):** Como se mencionó anteriormente todos los campos están separados por coma, pero en muchos casos la coma es parte del contenido de la celda, por lo cual el texto debe ir encerrado entre comillas dobles. En este caso la función bloqueante detecta el uso de la comilla doble, en caso de encontrar comilla simple en su lugar no permite la ejecución de las métricas, interrumpiendo el proceso. También cabe destacar, que en muchos casos la comilla doble puede ser parte del contenido, en el caso de no existir ninguna coma, la celda puede tener comillas dobles en cualquier parte de la celda, la comilla doble cuando se encuentra al inicio y al final del campo indica que el delimitador se encuentra en el interior del contenido, y si en este contenido también habría que representar el valor comilla doble (“”) se debe colocar doble comilla doble, por ejemplo: “JUAN, “”PEPE””,12” donde el contenido se interpretará como “PEPE” [CSV21].
- **EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS:** La siguiente validación es completamente estructural, en este caso se recorre ya todo el contenido del dataset, y comprueba que cada registro tenga la misma cantidad de campos como columnas haya en el dataset. En caso de faltar algún campo o bien sobrar alguno, se interrumpe el proceso y se descarta el dataset, con la notificación pertinente al usuario.
- **EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA):** El nombre CSV proviene del inglés, cuya traducción al español es “Valores Separados por Coma”, siendo este criterio un aspecto bloqueante para este tipo de archivos. Al leer el primer registro se detecta el uso de las comas para separar los campos, de no hallarse este delimitador se interrumpe el análisis del dataset dando aviso al usuario. Además, este es uno de los puntos que se indican como buenas prácticas [CSV21].



- **EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET:** Se verifican los títulos que deben estar contenidos en el primer registro. Esto es, para detectar los títulos de cada uno de los campos, éstos deben ser del tipo texto, por lo que se valida que dichos campos no estén vacíos, ni tampoco deben ser numéricos, ni fechas.
- **EL ARCHIVO POSEE NOMBRES DE TÍTULOS REPETIDOS:** El siguiente bloqueante está relacionado con los nombres de las columnas, los mismos no deben repetirse, entonces la comprobación que HEVDA realiza consiste en que cada campo sea único dentro del primer registro, de no cumplirse dicha condición, no proseguirá con las validaciones.

Al superar todas las validaciones de bloqueo, es decir, el archivo no posee tipos de bloqueos, se continúa con la ejecución del algoritmo que contiene el proceso de validación interno de cada métrica propuesta para evaluar la calidad del dataset.

En la Figura 5.1 sector A, se muestra la pantalla inicial de HEVDA que contiene la opción de selección del archivo y un botón “Analizar Dataset de Datos Abiertos” para iniciar la validación. Por otra parte, se observa una barra vertical en la parte izquierda con las categorías de métricas, críticas y no críticas, las que pueden ser desplegadas con un clic para acceder a sus correspondientes métricas establecidas. Esto se muestra en la Figura 5.1 sector B para métricas críticas y Figura 5.1 sector C para las métricas no críticas.



Figura 5.1 – Pantalla inicial de la herramienta HEVDA.

Una vez seleccionado el archivo a analizar, la herramienta mostrará un reporte para cada una de las métricas. A continuación, se presenta un ejemplo por cada una de éstas, con el fin de visualizar en pantalla como es la información arrojada.



5.2.1. Métricas

5.2.1.1. Tratamiento de Números Decimales

En la Figura 5.2, 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.

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).

⚠ VERIFICACIÓN DE NUMÉRICOS DECIMALES) (MEDICIÓN DE TIPOS DE DATOS)

Esta Métrica permite analizar los datos del tipo Decimal que no cumplen con las recomendaciones que se muestran a continuación, y además, permite obtener un análisis estimativo de Tipos de Campos del Conjunto de Datos

Dataset a procesado: operativo-detectar - copia - errordecimal3.csv

CANTIDAD DE DATOS QUE NO CUMPLEN CRITERIOS DE VALIDACIÓN

- ⊗ Hay 3 casos con números decimales mal cargados o separador decimal incorrecto con , (coma)
- ⊗ Los campos numéricos deben tener sólo 2 dígitos decimales.

[Clic para ver el Detalle de registros](#)

Análisis estimativo para el cálculo de los Tipos de Datos de los Campos del dataset. A continuación, la grilla con la estimación:

Columna	Tipo
r	Entero
t	Entero
Lat	Coordenada
Long	Coordenada
latitud	Coordenada
longitud	Coordenada
g	Decimal
data	Decimal
x	Coordenada
y	Coordenada

Figura 5.2 – Métricas Críticas: Pantalla de Verificación de Números Decimales.

En la Figura 5.3, 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	rg
2	6.988	La cantidad decimales es mayor a 2	data
3	11,43	El separador decimal es incorrecto	rg

Figura 5.3 – Detalle de los casos detectados para la verificación de números decimales.

5.2.1.2. Registros Duplicados

En la Figura 5.4, se muestra un ejemplo de la herramienta HEVDA, 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.

Por ejemplo: “Cantidad de casos detectados con duplicación de registros: 11 de 1400”.

En base a lo anterior, se presenta el porcentaje de registros duplicados afectados, siendo: “Porcentaje de registros duplicados: 0,79 %”.

Por otra parte, se podrá ver el detalle de los registros que se encuentran repetidos, desde la opción “Clic para ver el Detalle de registros repetidos”.

Datos Abiertos

Métricas Críticas ▾

Métricas No Críticas ▾

⚠ MEDICIÓN DE REGISTROS DUPLICADOS (MÉTRICA DE REDUNDANCIA)

Esta Métrica permite detectar y mostrar los registros repetidos en el Conjunto de Datos

Dataset a procesado: padron-de-operadores-organicos-certificados.csv

Cantidad de casos detectados con duplicación de registros: 11 de 1400

Porcentaje de registros duplicados: 0,79 %

[Clic para ver el Detalle de registros repetidos](#)

Figura 5.4 – Métricas Críticas: Pantalla de Verificación de Registros Duplicados.

Al realizar clic en el detalle de registros repetidos, se visualiza la pantalla que se muestra en la Figura 5.5, allí se puede observar una grilla de datos, con el número de registro afectado, y los números de registros duplicados para éste. Por otra parte, se agrega el detalle de toda la fila validada.





**MEDICIÓN DE REGISTROS DUPLICADOS
(MÉTRICA DE REDUNDANCIA)**

Registros encontrados

Registro afectado	Registros Duplicados	Detalle del Registro
6	12	32, ARGENTINA, 6, BUENOS AIRES, AZUL, INDEFINIDA, AGRICULTURA, "SOJA, ALPISTE", 1, Productores, 3, Letis S.A., AGRICOLA EL HORIZONTE S.A., EL RELINCHO
7	96 - 195	32, ARGENTINA, 6, BUENOS AIRES, AZUL, INDEFINIDA, AGRICULTURA, "SOJA, ALPISTE, TRIGO", 1, Productores, 3, Letis S.A., AGRICOLA EL HORIZONTE S.A., EL HORIZONTE
644	645	32, ARGENTINA, 54, MISIONES, SAN JAVIER, INDEFINIDA, AGRICULTURA, CAÑA DE AZUCAR, 1, Productores, 2, OIA S.A., VARGAS PACHECO JOSE CARLOS / DOMINGUEZ NETO ELIO / MACHADO BARCELLOS JOEL / MACHADO ARLINDO, IFAI - LOTE 28
1106	1107 - 1108	32, ARGENTINA, 46, LA RIOJA, CHILECITO, INDEFINIDO, SECADO DE FRUTAS, UVA PASA, 2, Elaboradores, 1, Argencert S.R.L., L.I.R.A. S.A., NC
1147	1148	32, ARGENTINA, 50, MENDOZA, LUJAN DE CUYO, INDEFINIDO, BODEGA VITIVINICOLA, VINO, 2, Elaboradores, 1, Argencert S.R.L., FINCA VON WIGSTEIN S.A., NC
1194	1195	32, ARGENTINA, 54, MISIONES, SANTO PIPO, INDEFINIDO, PROCESAMIENTO CULTIVOS INDUSTRIALES, YERBA MATE, 2, Elaboradores, 2, OIA S.A., KOTH ITATI MARIA EVANGELINA, NC
1238	1239	32, ARGENTINA, 62, RIO NEGRO, VIEDMA, INDEFINIDO, SECADO; PELADO; ENVASADO; ALMACENAMIENTO, "NUECES, NUEZ SECA PELADA, NUEZ SECA CON CASCARA", 2, Elaboradores, 3, Letis S.A., QUANTUM S.R.L., NC
1240	1241 - 1274	32, ARGENTINA, 62, RIO NEGRO, VALLE AZUL, INDEFINIDO, EMPAQUE Y FRIO, MANZANAS Y PERAS, 2, Elaboradores, 3, Letis S.A., SARALEGUY SANDRO SERGIO, NC

Figura 5.5 – Detalle de los casos detectados para la verificación de registros duplicados.

Por ejemplo, para el registro número 6, se detecta que se encuentra duplicado el registro 12, bajo el detalle del registro:

"32, ARGENTINA, 6, BUENOS AIRES, AZUL, INDEFINIDA, AGRICULTURA, "SOJA, ALPISTE", 1, Productores, 3, Letis S.A., AGRICOLA EL HORIZONTE S.A., EL RELINCHO".

Por otro lado, para el registro número 7, se detecta que se encuentra duplicado en más de un caso, siendo idéntico a los números de registros 96 y 195, bajo el detalle del registro:

"32, ARGENTINA, 6, BUENOS AIRES, AZUL, INDEFINIDA, AGRICULTURA, "SOJA, ALPISTE, TRIGO", 1, Productores, 3, Letis S.A., AGRICOLA EL HORIZONTE S.A., EL HORIZONTE".

5.2.1.3. Datos Faltantes y/o Completos

En la Figura 5.6, se muestra un ejemplo de la herramienta HEVDA, para la detección de datos faltantes y/o completos.

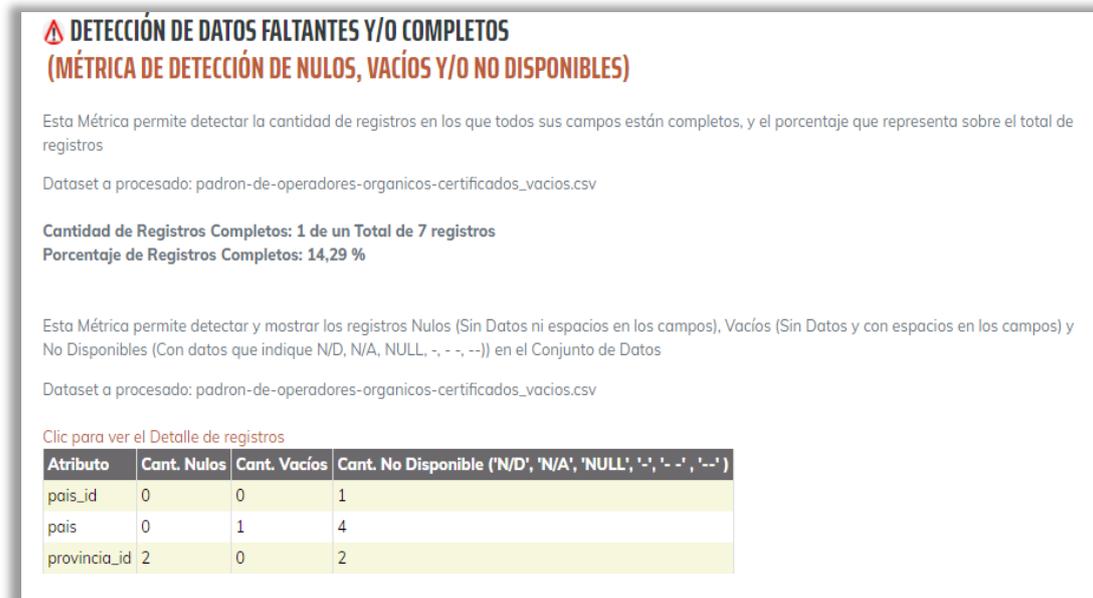


Figura 5.6 – Pantalla para la detección de datos faltantes y/o completos.

En la Figura 5.6, se puede ver la cantidad de registros completos y su correspondiente porcentaje, por ejemplo, se muestra:

- La cantidad de Registros Completos: “1 de un Total de 7 registros”.
- El porcentaje de Registros Completos: “14,29 %” (calculado sobre el total de registros).

El detalle de cada uno de estos registros identificados, se muestran en la Figura 5.7. Allí, se visualizan los números de registros, el nombre de la columna afectada, el tipo de clasificación (Nulo, Vacío o No Disponible) y el detalle del registro completo del dataset. Se observa una grilla de datos a modo de resumen, con las 3 (tres) clasificaciones propuestas discriminadas por columnas del dataset: Nulos, Vacíos y No disponibles.

Para este ejemplo:

- La clasificación de *Nulo*, tiene una cantidad de 2 casos, registrados para la columna “provincia_id”;
- Para la clasificación de *Vacíos*, se indica un caso para la columna “pais”;
- Para la clasificación de *No disponibles*, se contemplan los casos con datos que indiquen “N/D”, “N/A”, “NULL”, “-”, “- ” o bien “--”, para este grupo, existen 2 casos para la columna “provincia_id”, 4 casos para la columna “pais” y 1 caso para la columna “pais_id”.



**DETECCIÓN DE DATOS FALTANTES Y/O COMPLETOS
(MÉTRICA DE DETECCIÓN DE NULOS, VACÍOS Y/O NO DISPONIBLES)**

Registros encontrados

Nº Registro	Columna	Tipo	Registro Completo
2	pais	No Disponible	32,N/D,62.4599
3	pais	Vacío	32, ,
3	provincia_id	Nulo	32, ,
4	provincia_id	Nulo	32,ARGENTINA,
6	pais	No Disponible	566,NULL,aaaa
7	pais	No Disponible	566,N/A,-
7	provincia_id	No Disponible	566,N/A,-
8	pais_id	No Disponible	N/A,- ,--
8	pais	No Disponible	N/A,- ,--
8	provincia_id	No Disponible	N/A,- ,--

Figura 5.7 – Pantalla de detalle de casos para detección de datos faltantes.

Por ejemplo:

- Para el caso de la clasificación *Nulo*, se observan los registros 3 y 4, en la columna *provincia_id*.
- Para el caso de la clasificación *Vacío*, se observa el registro 3, en la columna “pais”.
- Para el caso de la clasificación *No Disponible*, se observa registro 8, en la columna “pais_id”.
- Para el caso de la clasificación *No Disponible*, se observan registros 2, 6, 7 y 8, en la columna “pais”.
- Para el caso de la clasificación *No Disponible*, se observan registros 7 y 8, en la columna “id_pais”.

5.2.1.4. Caracteres Inválidos

En la Figura 5.8, se muestra un ejemplo de la herramienta HEVDA, 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”. Por otra parte, se muestra una grilla de datos, con la cantidad de casos de caracteres especiales discriminados por columnas del dataset:

- Para la columna “nom_unimed” existen 2 casos.



- Para la columna “mes” existe 1 caso.
- Para la columna “cod_unimed” existe 1 caso.

⚠ VERIFICACIÓN DE CARACTERES INVÁLIDOS
(MÉTRICA DE CARACTERES ESPECIALES)

Esta Métrica permite identificar la cantidad de caracteres especiales y mostrarlos

Dataset a procesado: produccion-de-bioetanol-por-insumo-.csv

3 Columnas afectadas de un total de 9 del dataset

[Clic para ver el Detalle de los caracteres especiales](#)

Columna	Cant. de Casos
nom_unimed	2
mes	1
cod_unimed	1

Figura 5.8 – Pantalla de verificación de caracteres inválidos y la cantidad de casos por columnas.

La opción hacer “Clic para ver el Detalle de los caracteres especiales” muestra la grilla de la Figura 5.9. Esta es una grilla informativa que muestra los registros encontrados, indicando los números de registros, caracteres y columnas afectadas en el dataset.

Registros encontrados

Nro de Registro	Caracteres Afectados	Columna Afectada
3	~	cod_unimed
9	μ	nom_unimed
23	~	mes
26	~	nom_unimed

Figura 5.9 – Detalle de los registros encontrados para los caracteres inválidos.

5.2.1.5. Redundancia para el dominio de una columna

En la Figura 5.10, se muestra un ejemplo de la herramienta HEVDA, para la medición de redundancia en el dominio de valores de una columna. Para ello se visualiza la cantidad de columnas detectadas que tienen repetición en sus datos, sobre el total de columnas del dataset. Por ejemplo, se muestra la leyenda: “Se han encontrado: 10 columnas con campos duplicados de un total de 14 columnas”.



⚠ MEDICIÓN DE REDUNDANCIA EN EL DOMINIO DE VALORES DE UNA COLUMNA
(MÉTRICA DE VALORES DUPLICADOS EN COLUMNA)

Esta Métrica permite calcular la cantidad de valores repetidos en un mismo campo (dominio de valores)

Dataset a procesado: padron-de-operadores-organicos-certificados.csv

Se han encontrado: 10 columnas con campos duplicados de un total de 14 columnas

[Clic para el Detalle de la redundancia de dominios](#)

Figura 5.10– Pantalla de Medición de Redundancia en el dominio de valores de una columna.

La opción hacer “Clic para el Detalle de la redundancia de dominios”, muestra una pantalla, la cual se puede observar en la Figura 5.11.

Campos del DataSet:

Palabras del Campo:

Registros encontrados

Valor del Campo	Cantidad de Veces que se Repite	Nro Registro
CIUDAD AUTONOMA BUENOS AIRES	44	976 977 978 994 1030 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327

Figura 5.11– Detalle - Medición de la redundancia en el dominio de valores de una columna.

En dicha Figura 5.11, se muestra una lista desplegable con los campos del dataset que poseen repeticiones (para este ejemplo, 10 columnas), al realizar clic en una opción de la lista, se podrá visualizar una lista enlazada desplegable con las palabras, es decir, los valores que se repiten para dicho campo. A su vez, al seleccionar una opción de la segunda lista desplegable (palabra del campo), se muestra una grilla de datos con los registros encontrados que incluyen los números de registros del dataset y la cantidad de veces que se repite.

5.2.1.6. Redundancia entre campos de una misma fila

En la Figura 5.12, se muestra un ejemplo de la herramienta HEVDA, 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: “98 Registros afectados de un total de 1396 registros del dataset”.

Por otra parte, se muestra una grilla a modo de resumen con las combinaciones de columnas con datos redundantes detectados en sus campos, la cantidad de ocurrencias y el porcentaje que representa sobre el total de casos encontrados.



Figura 5.12– Pantalla de Medición de Redundancia entre campos de una misma fila.

Tomando a modo de ejemplo las dos primeras filas del resumen se puede observar que:

En la primera fila:

- Combinación de columnas: provincia, departamento.
- Cantidad de ocurrencias o apariciones de esta combinación: 61.
- Porcentaje de ocurrencia sobre el total de casos: 62,24 %

En la segunda fila:

- Combinación de columnas: razón social, establecimiento.
- Cantidad de ocurrencias o apariciones de esta combinación: 26.
- Porcentaje de ocurrencia sobre el total de casos: 26,53 %

La opción hacer “*Clic para ver el Detalle de Atributos Duplicados en Instancia Entidad*”, muestra una Pantalla, la cual se puede observar en la Figura 5.13. Allí, se tiene una lista desplegable con los campos del dataset que poseen repeticiones. Al realizar clic en una opción de la lista, automáticamente se carga una grilla con los registros encontrados que incluyen los números de registros y el dato redundante. Por ejemplo, en la figura, se listan los registros para la combinación de campos: razón social y establecimiento.



Campos del DataSet:

Registros encontrados

Nro de Registro	Dato redundante
103	PROVERBIO JUAN MARIANO
118	SASAKI NOOSAN SRL
223	FIDEICOMISO ARANDANOS DEL HEMISFERIO SUR
251	TIANA S.A.
304	OLIVARES S.A.
338	BUONO BENITO ARNALDO
364	FIDEICOMISO BAX
398	OJO DE VINO SA
468	CAVIGIOLI MARIO
515	FUNDACION ALBERTO ROTH
579	PAULUK TEODORO
608	SCHERF ALBERTO
671	GONZALEZ ALDO DONATO
674	JACOBO FRANCISCO JOSE
690	SAHORES S.A.
692	SANCHEZ GLADYS MABEL
702	AGRICO.AR S.A.
710	BANCO HUGO
718	BODEGA NOEMIA DE PATAGONIA SA
761	GARCIA JOAQUIN
792	MORRIS GLYNDWR LEONARDO
796	MUÑOZ MARIA LUISA
797	MUÑOZ MIGUEL
823	SANCHEZ PEDRO Y OMAR SH
836	VIZZOTTO RICARDO
837	VIZZOTTO ALBERTO

Figura 5.13– Detalle para la medición de la redundancia entre campos de una misma fila.

5.2.1.7. Detección de valores ID

En la Figura 5.14, 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. Es posible observar que se han encontrado 5 columnas que representan a ID (pais_id; id; provincia_id; categoria_id; Certificadora_id).

⚠ ESTIMACIÓN EN LA DETECCIÓN DE VALORES ID
(MÉTRICA DE ANÁLISIS DE ID)

Esta Métrica permite realizar una estimación sobre la cantidad de ID identificados en las columnas del dataset. Se analizan los nombres que contengan: 'id', 'id_' y/o '_id', tanto en mayúsculas como en minúsculas.

Dataset a procesado: padron-de-operadores-organicos-certificados - copia.csv

Título/Campo detectado en el análisis:
 pais_id;
 id;
 provincia_id;
 categoria_id;
 Certificadora_id;

Figura 5.14– Métrica de análisis de ID.



Muchas veces, estos códigos representan los identificadores que permitirán la interoperabilidad entre datasets, por lo que es importante detectarlos para lograr conocer el propósito de cada uno de éstos y así facilitar a un correcto análisis de los datos públicos abiertos.

5.2.1.8. Campos Triviales

En la Figura 5.15, se muestra un ejemplo de la herramienta HEVDA, para la verificación de campos triviales. Para ello se visualiza la cantidad de columnas afectadas sobre la cantidad total de columnas del dataset. Por ejemplo: “2 Columnas afectadas de un total de 14 del dataset”.

Por otra parte, se muestran las columnas afectadas con sus correspondientes valores detectados:

- Campo: *pais_id*; Valor del campo: 32.
- Campo: *pais*; Valor del campo: ARGENTINA.

Es decir, para el campo *pais_id* se encontró el valor “32”, y para el campo “*pais*” se detectó el valor “ARGENTINA” en todos sus datos.



Figura 5.15– Métrica de campos triviales.

5.3. Recolección de datos

En base a las funcionalidades presentadas sobre la herramienta HEVDA para las métricas propuestas, se realizó una selección de los conjuntos de datos abiertos que fueron validados con la aplicación programada. Para ello se tomaron como muestra, conjuntos de datos abiertos de 5 portales gubernamentales argentinos más relevantes:

- Portal de Datos Abiertos del Ministerio de Salud de Argentina [SAL21];
- Portal de Datos Abiertos de Diputados de la Argentina [DIP21];



- Portal de Datos Abiertos del Ministerio de Justicia y Derechos Humanos de la Argentina [MIN21];
- Portal de Datos Abiertos de la Ciudad de Buenos Aires [BAC21];
- Portal de Datos Abiertos de la República Argentina [AUD21].

Para cada portal se identificó los Temas/Categorías, y para cada uno de estos, se descargaron los datasets en formato abierto del tipo CSV para ser validados con la herramienta HEVDA. De cada portal se consideró un 25% del total de sus datasets a modo de muestra.

En la Tabla 5.1 se muestra la cantidad de datasets de cada uno de los portales gubernamentales y su correspondiente 25% como muestra considerada para la validación.

Tabla 5.1 – Cantidad de datasets para cada uno de los portales gubernamentales.

Portal de Datos Abiertos (relevados al 13 de enero 2021)	Cantidad Total de datasets	25% del total de datasets (cantidad)
Ministerio de Salud de Argentina [SAL21];	41	10
Diputados de la Argentina [DIP21];	29	7
Ministerio de Justicia y Derechos Humanos de la Argentina [MIN21];	61	15
Ciudad de Buenos Aires [BAC21]	398	99
República Argentina [AUD21]	1013	253
TOTAL de datasets a relevar		384

Para identificar el 25% de la muestra, en caso de tener un resultado con parte decimal mayor igual a 0,5 se toma un dataset adicional. Por ejemplo: De un portal de datos abiertos que tiene 61 datasets en total, su 25% de muestra es de 15,25 por lo que se consideraron 15 datasets. Para que la elección de los 15 dataset fuera aleatoria se contempló tomar una cantidad uniforme por categoría. Si el portal de datos públicos posee 10 Grupos/Temas/Categorías, se tomará en cuenta un dataset de cada Tema, más 1 dataset extra de 5 Temas con el fin de considerar los 15 datasets a modo de muestra.

Otra posibilidad que se presenta en la muestra es tener que tomar, por ejemplo, 5 datasets de cada Grupo, pero hay un Grupo que tiene sólo 2 datasets, por lo que se seleccionó más datasets de los siguientes grupos (según orden de aparición en el portal) para cubrir la muestra. Como se mostró en la Tabla 5.1, la muestra del 25% derivó en analizar una cantidad total de 384 datasets.

Para más información sobre los portales de datos abiertos gubernamentales seleccionados, se acompaña el [Anexo D: “Portales de datos abiertos gubernamentales”](#) con las capturas/evidencia de pantallas de los datasets totales por portal al 13 de enero del 2021.



5.4. Análisis de resultados

En esta sección se muestran todos los resultados obtenidos para las pruebas de los 384 datasets validados con la herramienta HEVDA. El análisis de resultados se presenta en distintas clasificaciones basadas en resultados generales, resultados discriminados por métricas críticas, no críticas y tipos de bloqueantes.

5.4.1. Análisis de las métricas bloqueantes

En esta sección se estudian los resultados de las métricas bloqueantes encontradas en la muestra de los 384 datasets. Dichos datasets bloqueantes, son descartados del análisis debido a la pertenencia de algún tipo de aspecto de incumplimiento por el cual no pueden ser tratados por la herramienta desarrollada.

En la Tabla 5.2, se muestra la cantidad total de datasets detectados como bloqueantes siendo 113 versus la cantidad de datasets no bloqueantes que son 271 sobre el total de los 384 datasets validados con la aplicación desarrollada.

Tabla 5.2 – Cantidad de datasets totales con bloqueantes.

TIPO DE BLOQUEANTE/ NO BLOQUEANTE	CANTIDAD DE DATASETS
DATASETS BLOQUEANTES	113
DATASETS NO BLOQUEANTES	271
TOTAL	384

Esta representación gráfica se muestra en la Figura 5.16, en la que el porcentaje de datasets bloqueantes representa un alto valor con un 29%, por lo que se puede sugerir detectar estos casos al preparar un conjunto de datos para evitar futuros inconvenientes en la interoperabilidad de los datasets gubernamentales públicos.

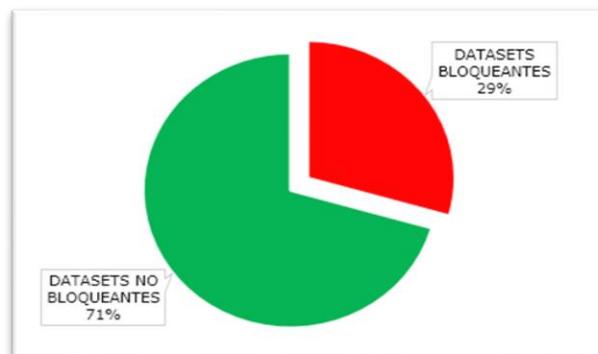


Figura 5.16 – Porcentaje de Tipos de bloqueantes y no bloqueantes detectados.



En la Tabla 5.3, se muestra el porcentaje que se establece para cada tipo de bloqueante entre los 113 datasets (29,43% de la muestra analizada), detectados con la herramienta HEVDA. Se observa que el primer lugar es para el Tipo 3 (“El archivo no cumple con el formato csv de separador (coma);”), que es el más representativo con más de la mitad de los tipos bloqueantes con el 57,52%. Seguido por el Tipo 1 (“El archivo posee doble carácter (comilla)”) con el 15,93%, luego por el Tipo 2 (“El archivo no cumple con la misma cantidad de columnas en cada uno de sus registros”) con el 14,16%, luego el Tipo 4 (“El archivo no posee una primera fila de títulos/nombres de las columnas del dataset”) con el 11,50% y en último lugar el Tipo 5 (“El archivo posee nombres de títulos repetidos”) con el 0,88%.

Tabla 5.3 – Cantidad de datasets por tipo de bloqueantes.

NÚMERO DE TIPO	TIPO DE BLOQUEANTE	CANTIDAD DE DATASETS
1	EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA);	15,93%
2	EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS;	14,16%
3	EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA);	57,52%.
4	EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET;	11,50%
5	EL ARCHIVO POSEE NOMBRES DE TÍTULOS REPETIDOS;	0,88%

Para más información sobre los resultados arrojados por la herramienta HEVDA de los 384 datasets testeados de los portales de datos abiertos gubernamentales seleccionados, se acompaña el [Anexo E: “Detalle de los resultados arrojados con HEVDA”](#), allí se muestran los nombres de datasets relevados junto al portal gubernamental que fue extraído, el cual es acompañado por los resultados de métricas de la 1 a la 8 inclusive (con estado cumple o no cumple).

5.4.2. Resultados generales de las métricas

En esta sección se muestra una comparativa de resultados arrojados en base al análisis de la cantidad de datasets que cumplen y no cumplen, segregadas por las 8 métricas propuestas en el capítulo 4.

La cantidad de datasets relevados es de 271, es decir el 70,57% de la muestra utilizada, ya que se encontraron 113 datasets, es decir el 29,43%, con características bloqueantes por lo que se descartan del análisis debido a la pertenencia de algún tipo de aspecto de incumplimiento por el cual no pueden ser tratados por la herramienta desarrollada. Estos bloqueantes, se explicaron en la sección 5.4.1.



En la Tabla 5.4, se observa el porcentaje en cada celda discriminadas por métricas que cumplen y las que no cumplen. La métrica que más se cumple en los datasets validados es la Métrica 2 con el 65,10% (registros duplicados), por otra parte, la métrica más incumplida es la Métrica 5 con el 55,73% (Redundancia en el dominio de valores de una columna).

Tabla 5.4 – Cantidad de datasets que cumplen y no cumplen por métrica.

MÉTRICAS	CUMPLE	NO CUMPLE
METRICA 1 Numéricos Decimales	54,43%	16,15%
METRICA 2 Registros Duplicados	65,10%	5,47%
METRICA 3 Datos Incompletos	40,36%	30,21%
METRICA 4 Caracteres Inválidos	33,85%	36,72%
METRICA 5 Redundancia en el dominio de valores de una columna	14,84%	55,73%
METRICA 6 Redundancia entre campos de una misma fila	44,01%	26,56%
METRICA 7 Detección de valores ID	47,40%	23,18%
METRICA 8 Campos Triviales	48,44%	22,14%

En la Figura 5.17, se muestra sólo la comparativa de métricas que cumple y no cumplen representada por gráficos de barra, en donde se visualiza en color verde la Métrica 2 (registros duplicados) con 250 datasets, es la que más se cumple, le sigue la Métrica 1 (validación del tipo de datos decimal) con 209 datasets, Métrica 8 (campos triviales) con 186 datasets y la Métrica 7 (detección de valores ID) con 182 datasets, siendo éstos los casos con menor dificultades encontradas. Otro de los aspectos que se observa, son las métricas que menos se cumplen con una barra en color rojo, como ser el caso de la Métrica 5 (redundancia de valores en el dominio de una columna) con 214 datasets, seguida por la Métrica 4 (caracteres inválidos) con 141 datasets y la Métrica 3 (datos incompletos) con 116 datasets, siendo éstos los 3 casos más relevantes en incumplimiento de calidad de datos.

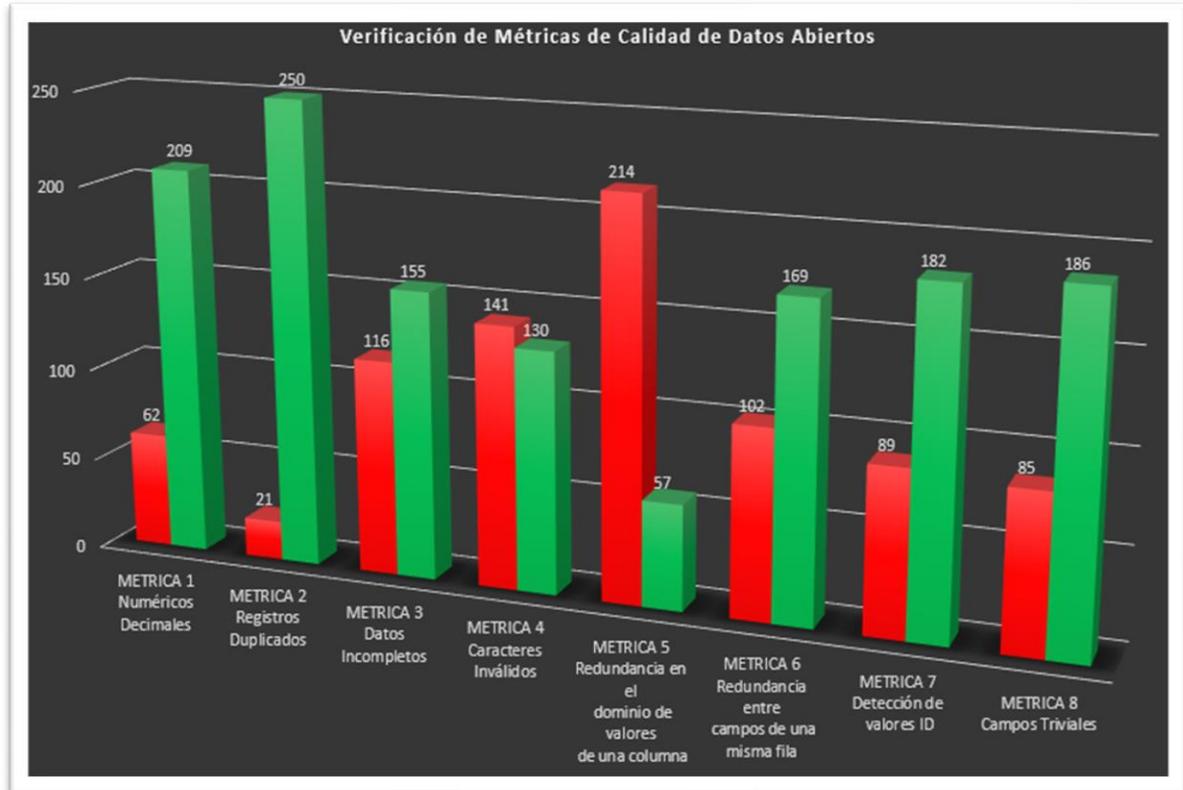


Figura 5.17 – Verificación de métricas de calidad de datos abiertos.

A continuación, se describen con mayor detalle, los resultados para las métricas críticas, no críticas y tipos de bloqueantes encontrados en las validaciones de los datasets.

En la Figura 5.18, se muestran los porcentajes de casos. Este estudio arroja que la métrica con mayores casos es la Métrica 5 (redundancia en el dominio de valores de una columna), representando un 25,78% entre todos los casos de incumplimiento, seguida de la Métrica 4 (caracteres inválidos) con un 16,99%, la Métrica 3 (datos incompletos) con un 13,98%, la Métrica 6 (redundancia entre campos de una misma fila) con el 12,29%, la Métrica 7 (detección de valores ID) con un 10,72%, la Métrica 8 (campos triviales) con un 10,24%, la Métrica 1 (validación del tipo de datos decimal) con un 7,47% y finalmente la Métrica 2 (registros duplicados) con un 2,53%. Esto concluye que las 3 métricas con mayor grado de incumplimiento son: Métrica 5, 4 y 3 respectivamente.

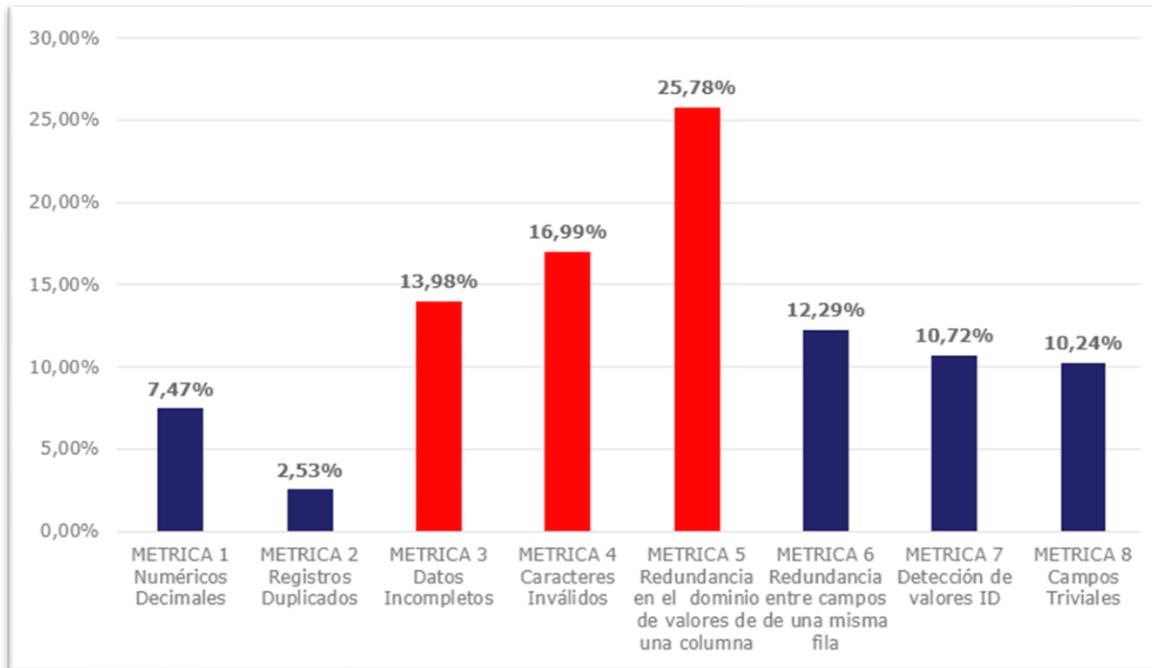


Figura 5.18 – Porcentaje de casos encontrados con incumplimiento de métricas.

Por otra parte, se incluye un gráfico en la Figura 5.19. que representa el estudio de las métricas críticas (de la 1 a la 4 inclusive) con el porcentaje que representa la cantidad de casos de incumplimiento para éstas. Es decir, de los 384 datasets validados, se encontraron 830 casos de incumplimiento de métricas de calidad de datos (siendo que un mismo dataset puede o no incumplir con más de una métrica), de los cuales se detectan 340 casos con incumplimiento de métricas críticas. Si bien no excede la mitad del total de los casos detectados (830 casos totales de incumplimiento), es un número alto bastante elevado.

En lo que respecta a métricas críticas, según la Figura 5.19, la métrica crítica más incumplida es la Métrica 4 (caracteres inválidos) con un 41,47%, seguida por la Métrica 3 (datos incompletos) con un 34,12%, luego la Métrica 1 (validación del tipo de dato decimal) con un 18,24% y finalmente la Métrica 2 (registros duplicados) con un 6,18%.

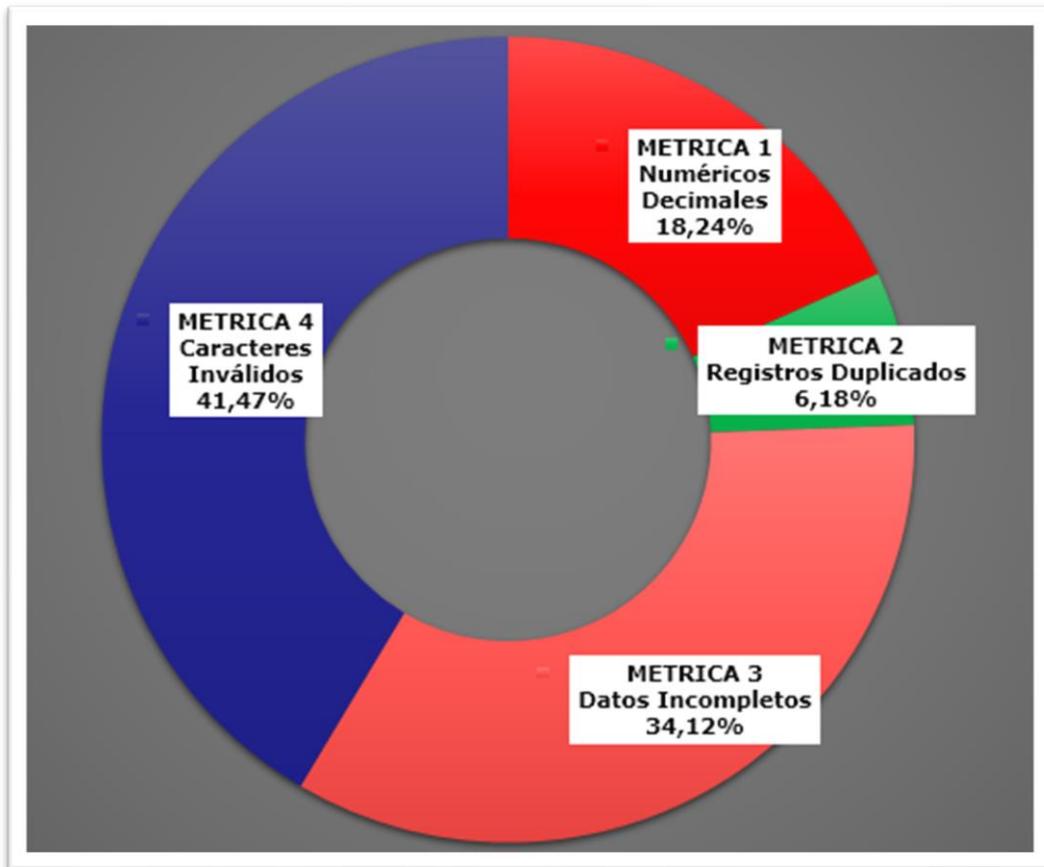


Figura 5.19 – Porcentaje de casos con incumplimiento de métricas críticas.

En la Figura 5.20. se muestra un gráfico con las métricas no críticas (de la 5 a la 8 inclusive) con el porcentaje que representa la cantidad de casos de incumplimiento para éstas. Es decir, de los 384 datasets validados, se encontraron 830 casos de incumplimiento de métricas de calidad de datos (siendo que un mismo dataset puede o no incumplir con más de una métrica), de los cuales se detectan 490 casos con incumplimiento de métricas no críticas. Este valor excede a más de la mitad del total de los casos detectados (830 casos totales de incumplimiento), por lo que, en cuestiones de incumplimiento, se encontraron más casos de métricas no críticas, que críticas.

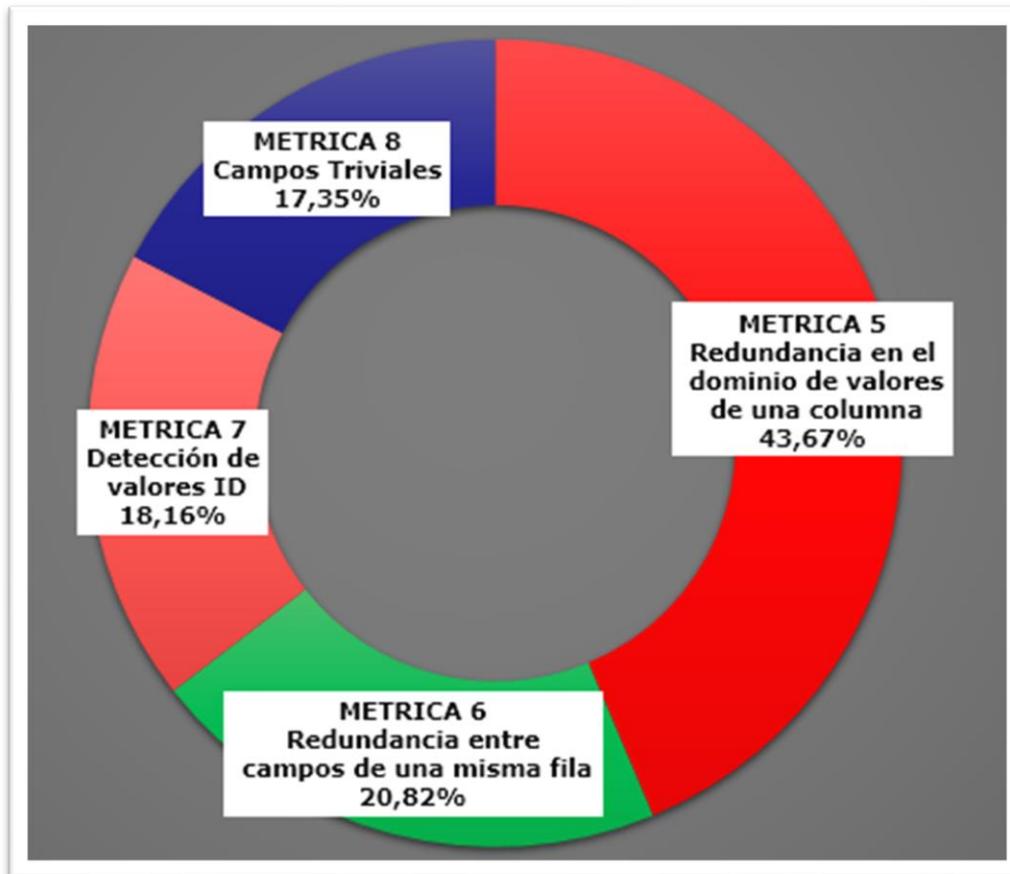


Figura 5.20 – Porcentaje de casos con incumplimiento de métricas no críticas.

Según la Figura 5.20., la métrica no crítica más incumplida es la Métrica 5 (redundancia en el dominio de valores de una columna) con un 43,67%, seguida por la Métrica 6 (redundancia entre campos de una misma fila) con un 20,82%, luego la Métrica 7 (detección de valores ID) con un 18,16% y finalmente la Métrica 8 (campos triviales) con un 17,35%.



Capítulo 6 – Validación

Para este capítulo, se realiza la validación de las métricas sugeridas en base a los distintos aspectos definidos que debe reunir un conjunto de métricas para la calidad de datasets abiertos. Esta definición parte del estudio realizado sobre la ISO/IEC 25012, el estándar universal de calidad de datos de dos capas, trabajos relevados, estudios realizados de desarrollo propio y la bibliografía relevada en esta tesis.

6.1. Introducción

La función principal de los datos abiertos gubernamentales es que sean reutilizarlos, 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” [CAD19]. 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” [PIÑ08].

El Portal Europeo de Datos [EUR21] 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, cuenta con más de 1.265.800 datasets. También ofrece información sobre el proceso de suministro de datos y los beneficios de reutilizarlos. En uno de sus informes [EUR19], 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*” [MIT02] 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 [DIM18], 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.

6.2. Aspectos de Validación

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* [ISO08], 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, 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* [CAI15], 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 [GCO20], 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* [DIM18]. Los criterios considerados para la comparativa de validación son: Completo, Exactitud, Consistencia, Integridad, Precisión y Representación.
- d) *Trabajos relevados* en el Capítulo 3 de la presente tesis y estudios de desarrollo propio [MAR20], [MAR21], [MAP20] y [MRO21].

En la Figura 6.1, se presentan los criterios que debe contener un conjunto de datos para que pueda ser de calidad e interoperable y que son definidos por el estándar universal de la calidad de los datos de 2 capas presentado en [CAI15]. Se visualiza el estándar universal en español extraído de



la guía de estándares de calidad e interoperabilidad de los datos abiertos del gobierno de Colombia [GES21]. Con un círculo se identifican los criterios considerados para la comparativa.

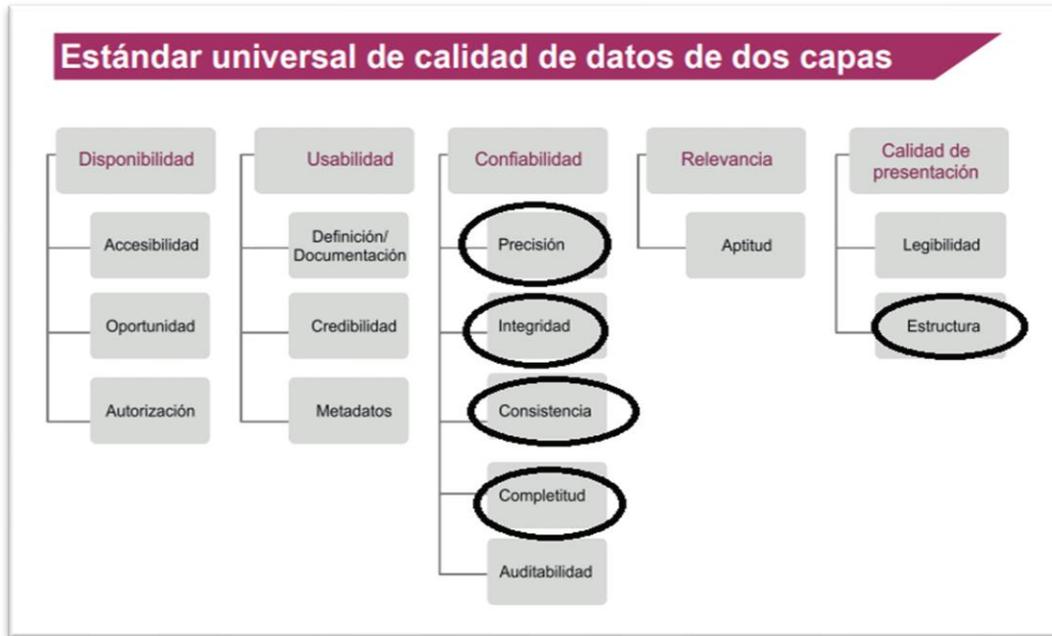


Figura 6.1 – Criterios seleccionados del Estándar universal de calidad de datos.

En la Figura 6.2, se presentan los criterios del modelo de calidad de datos definido por el estándar ISO/IEC 25012 [ISO08], en la imagen se muestran las 15 características que lo componen, y con un círculo se identifican los criterios considerados de la norma para la comparativa de validación de esta tesis.

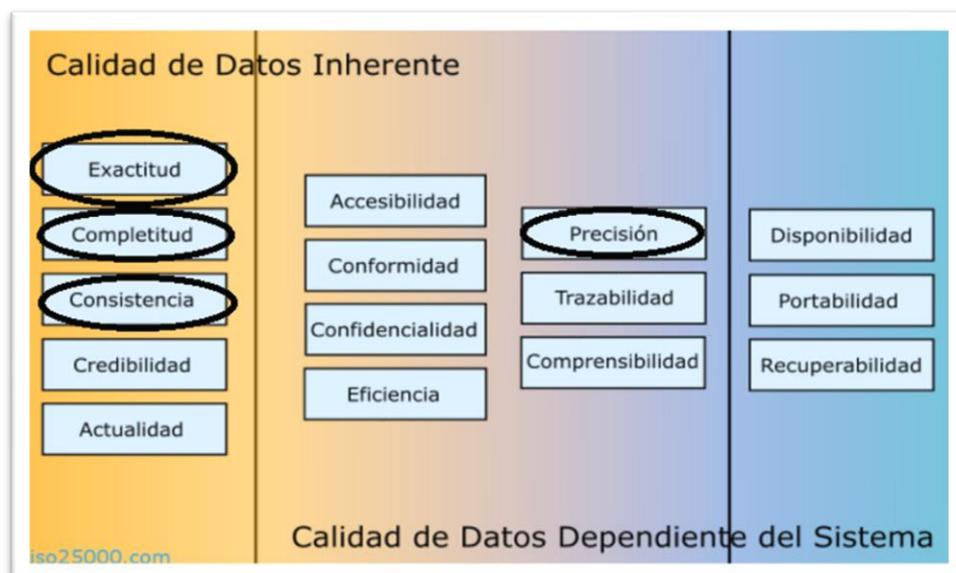


Figura 6.2 – Criterios seleccionados de la Norma ISO/IEC 25012.



En la Figura 6.3, se presentan las dimensiones de calidad de datos definido por Dan Myers [DIM18], y con un círculo se identifican los criterios considerados para la comparativa de validación de este capítulo.



Figura 6.3 – Criterios seleccionados de las dimensiones de calidad de datos.

A continuación, se presentan los criterios de comparación seleccionados y propuestos para las métricas sugeridas en la presente tesis.

6.2.1. 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 [ISO08], 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. A continuación, la Tabla 6.1 se muestran las métricas y aspectos que tienen relación directa con este aspecto de precisión, indicándose con una X.



Tabla 6.1 – Criterio de Precisión.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Precisión
1	CRÍTICAS: Números Decimales	X
2	CRÍTICAS: Registros Duplicados	
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	X
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	X
8	NO CRÍTICAS: Posibles campos triviales	X
Nro. Aspecto.	Aspectos generales	Criterio de Precisión
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	
2	Bloqueante: El archivo posee doble carácter de comilla	X
3	Bloqueante: Distinta cantidad de columnas en cada registro	
4	Bloqueante: Sin Formato CSV	
5	Bloqueante: Primera fila Sin nombres de columnas	
6	Bloqueante: Nombre de columnas repetidos	X

Se puede observar que en la Tabla 6.1, hay 8 métricas y/o aspectos que cumplen con el criterio de precisión representando el 57,14%.

6.2.2. 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 [ISO08]. A continuación, la Tabla 6.2 se muestran las métricas y aspectos que tienen relación directa con este aspecto de exactitud, indicándose con una X. Se puede observar que en la Tabla 6.2, se cumplen 8 aspectos de exactitud representando el 57,14%.

Tabla 6.2 – Criterio de Exactitud.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Exactitud
1	CRÍTICAS: Números Decimales	
2	CRÍTICAS: Registros Duplicados	X
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	X
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X



Nro. Met.	Métricas propuestas en la tesis	Criterio de Exactitud
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	
8	NO CRÍTICAS: Posibles campos triviales	X
Nro. Aspecto.	Aspectos generales	Criterio de Exactitud
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	
2	Bloqueante: El archivo posee doble carácter de comilla	X
3	Bloqueante: Distinta cantidad de columnas en cada registro	
4	Bloqueante: Sin Formato CSV	X
5	Bloqueante: Primera fila Sin nombres de columnas	
6	Bloqueante: Nombre de columnas repetidos	X

6.2.3. 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 [ISO08]. A continuación, la Tabla 6.3 se muestran las métricas y aspectos que tienen relación directa con este aspecto de completitud, indicándose con una X.

Tabla 6.3 – Criterio de Completitud.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Completitud
1	CRÍTICAS: Números Decimales	
2	CRÍTICAS: Registros Duplicados	
3	CRÍTICAS: Detección de Datos Incompletos	X
4	CRÍTICAS: Caracteres Inválidos	
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	
7	NO CRÍTICAS: Estimación de ID	
8	NO CRÍTICAS: Posibles campos triviales	
Nro. Aspecto.	Aspectos generales	Criterio de Completitud
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	
2	Bloqueante: El archivo posee doble carácter de comilla	
3	Bloqueante: Distinta cantidad de columnas en cada registro	
4	Bloqueante: Sin Formato CSV	
5	Bloqueante: Primera fila Sin nombres de columnas	X
6	Bloqueante: Nombre de columnas repetidos	



Se puede observar que en la Tabla 6.3, se cumplen 2 aspectos de completitud representando el 14,28%.

6.2.4. 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 [DIM18]. Su enfoque es sobre el hecho de conocer si los datos son relevantes de forma que se puedan utilizar. A continuación, la Tabla 6.4 se muestran las métricas y aspectos que tienen relación directa con este aspecto de integridad, indicándose con una X.

Tabla 6.4 – Criterio de Integridad.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Integridad
1	CRÍTICAS: Números Decimales	
2	CRÍTICAS: Registros Duplicados	X
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	X
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	X
8	NO CRÍTICAS: Posibles campos triviales	X
Nro. Aspecto.	Aspectos generales	Criterio de Integridad
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	X
2	Bloqueante: El archivo posee doble carácter de comilla	X
3	Bloqueante: Distinta cantidad de columnas en cada registro	X
4	Bloqueante: Sin Formato CSV	X
5	Bloqueante: Primera fila Sin nombres de columnas	X
6	Bloqueante: Nombre de columnas repetidos	X

Se puede observar que en la Tabla 6.4, se cumplen 12 aspectos de integridad representando el 85,71%.

6.2.5. 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 [DIM18]. Según [ISO08], 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. A continuación, la Tabla 6.5 se muestran las métricas y aspectos que tienen relación directa con este aspecto de consistencia, indicándose con una X.



Tabla 6.5 – Criterio de Consistencia.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Consistencia
1	CRÍTICAS: Números Decimales	X
2	CRÍTICAS: Registros Duplicados	X
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	X
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	X
8	NO CRÍTICAS: Posibles campos triviales	X
Nro. Aspecto.	Aspectos generales	Criterio de Consistencia
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	X
2	Bloqueante: El archivo posee doble carácter de comilla	X
3	Bloqueante: Distinta cantidad de columnas en cada registro	X
4	Bloqueante: Sin Formato CSV	X
5	Bloqueante: Primera fila Sin nombres de columnas	X
6	Bloqueante: Nombre de columnas repetidos	X

Se puede observar que en la Tabla 6.5, se cumplen 13 aspectos de consistencia representando el 92,85%.

6.2.6. 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. A continuación, la Tabla 6.6 se muestran las métricas y aspectos que tienen relación directa con este aspecto de consistencia, indicándose con una X.

Tabla 6.6 – Criterio de relación de valores.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Relación
1	CRÍTICAS: Números Decimales	
2	CRÍTICAS: Registros Duplicados	X
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	
8	NO CRÍTICAS: Posibles campos triviales	X



Nro. Aspecto.	Aspectos generales	Criterio de Relación
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	
2	Bloqueante: El archivo posee doble carácter de comilla	
3	Bloqueante: Distinta cantidad de columnas en cada registro	X
4	Bloqueante: Sin Formato CSV	
5	Bloqueante: Primera fila Sin nombres de columnas	
6	Bloqueante: Nombre de columnas repetidos	X

Se puede observar que en la Tabla 6.6, se cumplen 6 aspectos de relación de datos representando el 42,86%.

6.2.7. 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 [DIM18], 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. A continuación, la Tabla 6.7 se muestran las métricas y aspectos que tienen relación directa con este aspecto de consistencia, indicándose con una X.

Tabla 6.7 – Criterio de estructura/Representación.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Estructura
1	CRÍTICAS: Números Decimales	X
2	CRÍTICAS: Registros Duplicados	
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	
7	NO CRÍTICAS: Estimación de ID	
8	NO CRÍTICAS: Posibles campos triviales	
Nro. Aspecto.	Aspectos generales	Criterio de Estructura
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	X
2	Bloqueante: El archivo posee doble carácter de comilla	X
3	Bloqueante: Distinta cantidad de columnas en cada registro	X
4	Bloqueante: Sin Formato CSV	X
5	Bloqueante: Primera fila Sin nombres de columnas	X
6	Bloqueante: Nombre de columnas repetidos	X



Se puede observar que en la Tabla 6.7, se cumplen 7 aspectos de estructura de datos representando el 50%.

6.2.8. Redundancia

El criterio de redundancia permite analizar si se tiene la misma información 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, la Tabla 6.8 se muestran las métricas y aspectos que tienen relación directa con este aspecto de consistencia, indicándose con una X.

Tabla 6.8 – Criterio de redundancia.

Nro. Met.	Métricas propuestas en la tesis	Criterio de Redundancia
1	CRÍTICAS: Números Decimales	
2	CRÍTICAS: Registros Duplicados	X
3	CRÍTICAS: Detección de Datos Incompletos	
4	CRÍTICAS: Caracteres Inválidos	
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X
7	NO CRÍTICAS: Estimación de ID	
8	NO CRÍTICAS: Posibles campos triviales	X
Nro. Aspecto.	Aspectos generales	Criterio de Redundancia
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.	
2	Bloqueante: El archivo posee doble carácter de comilla	
3	Bloqueante: Distinta cantidad de columnas en cada registro	
4	Bloqueante: Sin Formato CSV	
5	Bloqueante: Primera fila Sin nombres de columnas	
6	Bloqueante: Nombre de columnas repetidos	X

Se puede observar que en la Tabla 6.8, se cumplen 5 aspectos de redundancia de datos representando el 35,71%.

Para más información sobre la visión del enfoque utilizado para la deliberación de cada criterio de validación, se presenta el [“Anexo F: Detalle de los criterios de Validación”](#).



6.3. Análisis de los criterios aplicados a las métricas sugeridas

En la sección 6.2 se presentaron los criterios de validación seleccionados para las métricas sugeridas en la presente tesis. Como resultado de la sección anterior, se muestra la Tabla 6.9 con el resumen del análisis, identificando que todas las métricas como los aspectos generales influyen en uno o más criterios.

Tabla 6.9 – Criterios de Validación.

Nro. Métrica	Métricas propuestas	Precisión	Exactitud	Complejidad	Integridad	Consistencia	Relación	Estructura	Redundancia
1	CRÍTICAS: Números Decimales	X				X		X	
2	CRÍTICAS: Registros Duplicados		X		X	X	X		X
3	CRÍTICAS: Detección de Datos Incompletos			X					
4	CRÍTICAS: Caracteres Inválidos	X	X		X	X			
5	NO CRÍTICAS: Redundancia en el Dominio de una columna	X	X		X	X	X		X
6	NO CRÍTICAS: Redundancia entre valores de Columnas de un mismo Registro	X	X		X	X	X		X
7	NO CRÍTICAS: Estimación de ID	X			X	X			
8	NO CRÍTICAS: Posibles campos triviales	X	X		X	X	X		X
Nro. Aspecto	Aspectos generales	Precisión	Exactitud	Complejidad	Integridad	Consistencia	Relación	Estructura	Redundancia
1	Aspecto de Estructura: Columnas con inicio/fin con espacios; Cant filas y columnas; UTF-8.				X	X		X	
2	Bloqueante: El archivo posee doble carácter de comilla	X	X		X	X		X	
3	Bloqueante: Distinta cantidad de columnas en cada registro				X	X	X	X	
4	Bloqueante: Sin Formato CSV		X		X	X		X	
5	Bloqueante: Primera fila Sin nombres de columnas			X	X	X		X	
6	Bloqueante: Nombre de columnas repetidos	X	X		X	X	X	X	X

En la Figura 6.4, se observa un estudio completo de los criterios de validación y su correspondiente porcentaje de implicancia en cada métrica y aspecto general. Allí se puede observar que el criterio más representativo para validar la calidad de datos abiertos es el aspecto de Consistencia con un



92,85%, seguida por el aspecto de Integridad con un 85,71%. Luego, con un doble empate en porcentaje, el criterio de Precisión y Exactitud con el 57,14%. A éstos les sigue el criterio de Estructurales/Representación con el 50%, continuando con el 42,86% para el criterio de Relación entre valores de campos. Finalmente, concluye el criterio Redundancia con el 35,71%.

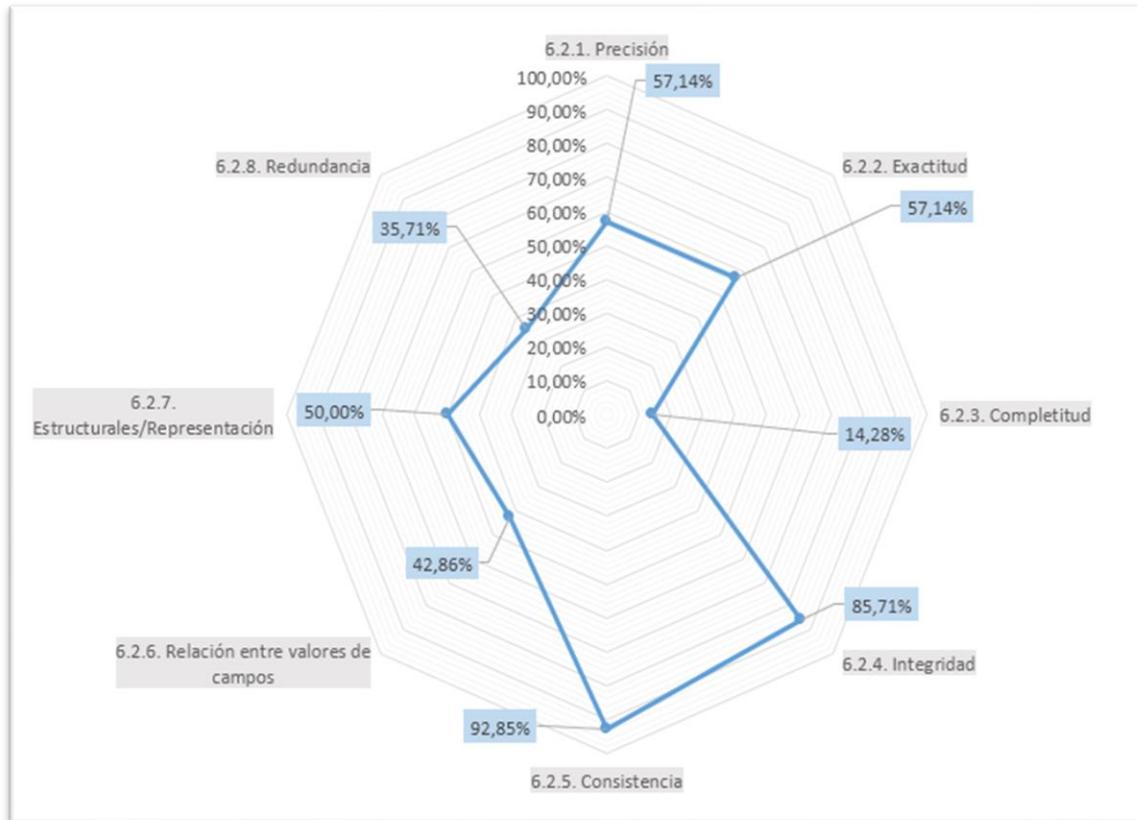


Figura 6.4 – Resumen de criterios de validación.

Esto conlleva a que las métricas propuestas se enfocan en un mayor porcentaje en métricas de validación orientadas a la “Consistencia”, como así también a la “Integridad” de los valores que contienen los datasets, por lo que su importancia radica en la coherencia general de los datos y el valor fiable que se encuentra almacenado. Por ejemplo, un conjunto de datos que contiene información sobre las personas internadas actualmente en un hospital determinado, se la puede considerar *inconsistente* si el recuento de las personas es mayor que el número de camas registradas. En lo que respecta a la *Integridad*, reúne un conjunto de características de calidad de datos mayores, relacionando más parámetros, como ser la precisión, fiabilidad y coherencia de los datos.

En el siguiente capítulo se muestran las conclusiones de la presente tesis, como así también los artículos presentados y trabajos futuros.



Capítulo 7 – Conclusiones y Trabajo Futuro

En este capítulo se presenta las conclusiones alcanzadas en base al estudio realizado, las contribuciones y reflexiones finales, las líneas de trabajo futuro y finalmente, las presentaciones a revistas y congresos que dan soporte al trabajo desarrollado en la presente tesis.

7.1. Reflexiones Finales

El paradigma de Gobierno Abierto, hoy por hoy desempeña un rol fundamental entre la conexión de distintos organismos gubernamentales y los ciudadanos. Gracias a este nuevo concepto es posible una mejor transparencia en base al tratamiento de rendición de cuentas por parte de entidades estatales, como así también, de una participación más comprometida con los ciudadanos de un país.

Según se muestra en este estudio, el papel que desempeñan las TICs es fundamental para la gestión de los datos abiertos públicos, por lo que es necesario verificar y realizar un seguimiento constante de todos los conjuntos de datos que son compartidos a través de los portales de sitios web gubernamentales, ya que son las fuentes de distribución para generar nuevo conocimiento como valor agregado a la comunidad.

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.

Los resultados de los sondeos realizados 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.

La importancia de 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. Cabe destacar que las métricas de calidad propuestas por la herramienta HEVDA que se desarrolló en este trabajo, permiten mitigar posibles errores en el tratamiento de las fuentes de datos, además, este es un punto positivo para la colaboración en la interoperabilidad del software, para que los datos abiertos puedan ser reutilizados. Un punto interesante para tener en cuenta es que la herramienta de validación HEVDA podría utilizarse en los portales abiertos de las diferentes agencias del gobierno estatal, con el fin de brindar un estudio que ayude a aquellas organizaciones que quieran poner a disposición sus datos públicos abiertos. Esto generaría conciencia sobre la importancia del valor de los datos para realizar estudios válidos, además de tener en cuenta que las organizaciones podrán validar sus conjuntos de datos antes de ser publicados en sus sitios web y así, poder mitigar o evitar ciertos errores en su contenido.

Los datos abiertos, provistos por entidades gubernamentales, tienen un sello de calidad y confiabilidad, que tiene un papel clave en la ciudadanía, ya que tanto los ciudadanos como las organizaciones, pueden monitorear activamente las políticas públicas involucradas. Por otro lado, otro punto discutido en esta investigación es que si bien existen portales de datos abiertos que mantienen actualizados sus conjuntos de datos, esto no asegura que la información obtenida sea completa y correcta. Como parte de las limitaciones de este trabajo, se indica que la herramienta de validación solo permite la lectura y verificación de archivos de datos abiertos del tipo CSV, por lo que sería interesante ampliar las posibilidades en el tratamiento de formatos más abiertos a futuro.

La presente tesis menciona varias fuentes de autores, en las que se evidencia la importancia de tener presente la calidad de los datos en aspectos de fuentes de datos abiertos gubernamentales. Otro de los puntos analizados fue el contenido de los distintos sitios web oficiales de varios países, en los que los aspectos de Gobierno Abierto están cada vez más presentes. Por otro lado, se analizaron trabajos que ofrecen estándares de medición y buenas prácticas para la evaluación de conjuntos de datos en referencia a la calidad. Por otro lado, es necesario realizar periódicamente un análisis en profundidad del estado de los países que se guían por el enfoque de Gobierno Abierto, ya que esto permitirá una mejor concienciación entre los ciudadanos y permitirá conocer el estado actual de la calidad de datos ofrecidos a personas de todo el mundo.

Como se analizó en secciones anteriores, existen varias carencias en aspectos de calidad de los datos que se brindan en los datos abiertos, por lo que compartir buenas prácticas, guías de aprendizaje, foros o herramientas que permitan un mejor análisis y tratamiento de los datos es un beneficio para el ciudadano que desea acceder a la información pública.

La propuesta de la herramienta HEVDA permite validar rápidamente el estado de "salud" de un conjunto de datos y analizar algunas de las métricas propuestas en aspectos de calidad, esto ayudará a saber si un conjunto de datos debe ser modificado o si está en condiciones de servir como base y analizar aspectos más profundos de los problemas de interoperabilidad del software para obtener valor agregado.

La contribución de este trabajo proporciona una herramienta que ofrece un análisis rápido de varios aspectos de calidad junto con una serie de métricas críticas y no críticas.



7.2. Ejecución de los Objetivos

Los datos abiertos gubernamentales son un recurso público que se encuentra en pleno auge. Los gobiernos abiertos generan valor a través de éstos, favoreciendo así a la transparencia y gestión participativa entre los ciudadanos, y, además, fomenta la mejora en la eficiencia y eficacia de los servicios brindados por administración pública, como así también, promueve una oportuna comprensión del contexto actual a través de las fuentes de datos combinadas.

El objetivo general de la presente tesis doctoral se orientó en proponer una serie de métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales. A raíz de esto, en el capítulo 1, se plantearon objetivos más específicos para cumplir con el mismo. A continuación, se presentan los resultados obtenidos en base a los objetivos específicos:

- *Objetivo Específico 1: Estudiar el alcance de los conceptos implicados en este contexto de estudio:*

En el Capítulo 2 se presentaron los conceptos más relevantes de la tesis a través de un marco teórico para el paradigma de gobierno abierto, con las definiciones e implicancias de datos abiertos y públicos. Además, se indicó el marco normativo y legal que se presenta para este contexto, como así también la importancia del acceso a la información pública. A modo complementario, se expusieron las definiciones de transparencia activa y pasiva.

También fueron mostrados los niveles de apertura de los datos abiertos con las consideraciones a tener en cuenta para disponibilizar los datos en los portales abiertos gubernamentales. Otro de los puntos estudiados, fue el estado situación sobre los datasets a nivel internacional, a través de un mapa de impacto mundial y una clasificación segmentada por los distintos gobiernos por medio del barómetro [ODB21]. Luego, se relevó la situación actual de Argentina y la adaptación de los catálogos de datos en los diversos sitios web estatales.

Finalmente, se culmina el capítulo con el relevamiento de los principios internacionales propuestos por el Open Data Charter, Carta Internacional de Datos Abiertos [ODC21] y el Open Government Data, Datos de Gobierno Abierto [OGD21], en donde para ambas entidades, se indicaron los estándares de buenas prácticas para la publicación y difusión de los datasets.

Otro aspecto esencial es la calidad de los portales de datos, es por ello que se relevó una serie de medidas cuantitativas a tener en cuenta sobre los aspectos referidos a la calidad de los datos. Esto se detalló mediante el modelo de calidad definido por el estándar ISO/IEC 25012 [ISO12].

- *Objetivo Específico 2: Analizar los trabajos relacionados en la temática de Gobierno Abierto relativos al tratamiento de datos abiertos y públicos; Objetivo Específico 3: Relevar los trabajos relacionados en cuanto a los aspectos de calidad de datos abiertos y públicos; Objetivo Específico 4: Analizar los criterios de los portales de datos abiertos para efectuar la publicación de éstos; Objetivo Específico 5: Relevar los distintos tipos y formatos*



disponibles de datos abiertos que existen en los dataset actuales de los sitios más relevantes y gubernamentales de Argentina:

Estos objetivos fueron estudiados en el Capítulo 3, en el desarrollo de un relevamiento de trabajos relacionados y situación actual. Algunos criterios de análisis fueron: la construcción de los repositorios de datos abiertos y sus características especiales, los aspectos que fueron considerados en la evaluación para llevar a cabo una correcta reutilización de los datasets, por otra parte, se indicaron los elementos protagonistas en el entorno del crowdsourcing y su interacción con los ciudadanos en gobierno abierto, como así también, un breve estudio de los tipos de fuentes de datos más utilizados.

Además, se analizaron trabajos respectivos sobre las consideraciones de los diversos autores en base a la calidad del dato público. Esto se acompañó con los criterios más relevantes en lo relativo a guías de buenas prácticas para el tratamiento, seguimiento y gestión del dato abierto público.

Finalmente, se expusieron las herramientas de validación de datasets más notables que se encuentran en la actualidad, como en versión beta.

- Objetivo Específico 6: Analizar las falencias de los datasets disponibles en los sitios de Gobierno; Objetivo Específico 7: Establecer criterios estándar de calidad de datos; Objetivo Específico 8: Proponer métricas críticas y no críticas de calidad de datos abiertos para validar diversos datasets.

Estos objetivos fueron trabajados en el Capítulo 4, en el que se indicaron una serie de inconvenientes detectados en los datasets de los portales de datos abiertos, como así también cuestiones que podrían ser contempladas para una mejor calidad en la presentación del dato almacenado en los archivos abiertos. En base a las falencias encontradas, se propusieron métricas críticas y no críticas basadas en las buenas prácticas de algunas publicaciones de los datos abiertos [TIP21], como así también, se efectuó un análisis sobre el formato más utilizado en datasets, y recomendaciones basadas mayormente en las especificaciones de la W3C [W3C15], y, además, en la experiencia de datasets relevados de diversos sitios gubernamentales de portales gubernamentales [MAR20], [MAR21], [MAP20] y [MRO21].

En dicho capítulo, no sólo se mencionaron las métricas propuestas, sino que, además, se explicaron y ejemplificaron cada una de éstas, brindando una justificación específica sobre el aporte sustancial que presenta en la calidad del dato y su importancia.

- Objetivo Específico 9: Desarrollar una herramienta de validación de datasets gubernamentales;

Este objetivo se trabajó en el Capítulo 5, en el que se presentó la herramienta HEVDA programada que permite la validación de las distintas métricas sugeridas para un conjunto de datos abiertos en formatos CSV (valores separados por comas). Se mostró la herramienta y sus funciones principales con ejemplos para cada una de las métricas, capturas de pantalla de la aplicación propuesta, y sus correspondientes explicaciones.



Además, se acompañó un detalle de los aspectos técnicos más relevantes, como así también los aspectos funcionales de ésta.

- Objetivo Específico 10: Definir y recolectar la muestra de datasets para ser testeados por la herramienta propuesta como validadora de calidad;

En la sección 5.3 del Capítulo 5, se identificó la selección de los conjuntos de datos abiertos que fueron validados con la aplicación programada. Para ello se tomaron como muestra 5 portales gubernamentales argentinos de datos abiertos más relevantes [SAL21], [DIP21], [MIN21], [BAC21] y [AUD21]. También se explicó la metodología de elección de la muestra de datos y los datasets escogidos de cada portal.

- Objetivo Específico 11: Definir un análisis cuantitativo en la medición de las métricas propuestas a través de la herramienta desarrollada; Objetivo Específico 12: Analizar los resultados arrojados con la herramienta programada para las métricas de calidad propuestas:

En la sección 5.4 del Capítulo 5, se analizaron los resultados obtenidos para el relevamiento de los 384 datasets validados con la herramienta HEVDA.

Se presentaron los tipos de bloqueantes que fueron encontrados en la muestra trabajada, como así también, la explicación de cada uno de éstos y las cantidades detectadas. El estudio arrojó resultados que fueron segmentados en aspectos generales, identificando parámetros de criterios cumplidos y no cumplidos, métricas generales, gráficos comparativos, como así también porcentaje de casos incumplidos en métricas críticas y no críticas.

- Objetivo Específico 13: Validar las métricas sugeridas en base a criterios estándares.

Este objetivo fue tratado en el Capítulo 6, en el que se realizó una validación de las métricas sugeridas en base a los distintos aspectos definidos que debe reunir un conjunto de métricas para la calidad de datasets abiertos. Esta definición parte del estudio realizado sobre la ISO/IEC 25012, el estándar universal de calidad de datos de dos capas, estudios de investigación propios y la bibliografía relevada en esta tesis.

Con los resultados obtenidos en el estudio realizado de la presente tesis, se puede concluir que sí 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 y sencilla.

7.3. Líneas Futuras de Investigación

Algunos de los aspectos que pueden ser estudiados como parte de los trabajos futuros de la presente tesis son los siguientes:

- Ampliar el relevamiento y estudio de nuevos aspectos que surjan en base a las nuevas necesidades de las plataformas de portales públicos de datos abiertos orientados a la medición de calidad en los datasets.



- Continuar trabajando en la ampliación de la herramienta de validación HEVDA, incorporando más criterios y principios mencionados en este trabajo por las distintas organizaciones mundiales que inculcan la cultura de gobierno abierto.
- Incorporar a la herramienta desarrollada más tipos de formatos abiertos, como ser: XML, JSON, etc. lo que llevará a establecer nuevos aspectos de control en las estructuras de los formatos que serán implementados.
- Analizar las opciones de aplicaciones gráficas que podrían ser utilizadas con los datasets para ser embebidos en el código fuente de la herramienta HEVDA.
- Incorporar una herramienta de gráficos estadísticos con el análisis de las métricas críticas y no críticas.
- Analizar los datasets orientados a la geolocalización para el tipo de contexto: Coordenadas de longitud y latitud, formatos geoespaciales, archivos del tipo WKT (puntos de coordenadas), SHP (datos geográficos), etc.
- Efectuar un estudio detallado de los datasets geoespaciales para proponer métricas de calidad de datos para este entorno de trabajo.
- Analizar la posibilidad de incorporar la utilización del prototipo HEVDA, en una entidad gubernamental.

7.4. Publicaciones en Revistas y Congresos

Dentro de las líneas de investigación que fueron efectuadas, surgieron algunas de las publicaciones que inicialmente se orientaron al concepto de Gobierno Abierto, luego, se realizó una especialización dentro del estudio sobre el paradigma de datos abiertos públicos gubernamentales y finalmente, se continuó sobre el estudio detallado de la calidad de los datos abiertos públicos que son ofrecidos en los portales gubernamentales, mejor conocidos como datasets. A continuación, se muestran las publicaciones respectivas que fueron elaboradas para los congresos nacionales e internacionales, en el contexto de la presente tesis. Éstas son:

- **Publicación N°1:**
 - **Fecha:** 8 al 12 de octubre de 2018.
 - **Título:** “11729 - Análisis Del Diseño Adaptativo En Sitios Web Gubernamentales”.
 - **Congreso:** XXIV Congreso Argentino de Ciencias de la Computación (CACIC 2018).
 - **Lugar:** Universidad Nacional del Centro de la Provincia de Buenos Aires (UNCPBA). Tandil.
 - **Autores:** Roxana Martínez; Rocío Andrea Rodríguez; Pablo Vera.
 - **Relación con la Tesis:** Este artículo realiza un relevamiento sobre las tecnologías utilizadas en algunos de los sitios web gubernamentales más relevantes del sector público, con el fin de conocer las técnicas de usabilidad y así comprender si se logra una mejor comunicación entre Estado y sus ciudadanos en base a la información



brindada. Lo anteriormente dicho, se analiza desde un punto de vista del diseño web adaptativo, detectándose así, que problemas son los que dificultan la usabilidad de las personas en lo que respecta a la información publicada, siendo este uno de los elementos fundamentales en Gobierno Abierto. A modo de aporte complementario, este estudio puede ser replicado por otros colegas a fin de validar otros sitios web estatales para evaluar su estado general.

▪ **Publicación N°2:**

- **Fecha:** 14 al 18 de octubre de 2019.
- **Título:** “Análisis de técnicas de raspado de datos en la web – aplicado al portal del Estado Nacional Argentino”.
- **Congreso:** Congreso Argentino de Ciencias de la Computación (CACIC 2019).
- **Lugar:** Universidad Nacional de Río Cuarto. Córdoba, Argentina.
- **Autores:** Martínez, Roxana; Rodríguez, Rocío; Vera, Pablo; Parkinson, Christian.
- **Relación con la Tesis:** Continuando con el análisis sobre la importancia de la tecnología en sitios web gubernamentales, este artículo se enfoca en el estudio del dato, es decir, entender la importancia de éstos como instrumento fundamental para el avance en la tecnología estatal. Garantizar la calidad de los datos no es sencillo, por lo que estudiar los procesos de recolección de estos, puede ayudar a un buen estudio. Este trabajo propone distintos aspectos para tener en cuenta para el tratamiento de los datos y lograr que sean más precisos, actualizados y completos en técnicas de captura de información. Para ello, se analizan distintas herramientas de raspado de datos (web scraping) que existen en el mercado y se estudian las funciones principales de utilización. Se analizan casos de estudios sobre sitios web del Estado Nacional Argentino, con el fin de extraer datos públicos y analizar así, el procedimiento de estos a una fuente de datos para su posterior tratamiento.

▪ **Publicación N°3:**

- **Fecha:** 21 y 22 de noviembre de 2019.
- **Título:** “*Criterios para el análisis preliminar de datos extraídos mediante web scraping*”.
- **Congreso:** III Congreso Internacional de Ciencias de la Computación y Sistemas de Información (CICCSI 2019).
- **Lugar:** Universidad Champagnat. Mendoza, Argentina.
- **Autores:** Martínez, Roxana; Parkinson, Christian; Rodríguez, Rocío; Vera, Pablo.
- **Relación con la Tesis:** Este trabajo analiza técnicas que permiten una mejor validación y tratamiento de buenas prácticas en lo que respecta a la calidad de los datos extraídos a través de métodos de raspado de sitios web gubernamentales. Para tener calidad en la información obtenida desde un sitio web, es importante el enfoque de tipos de validaciones de datos, ya que algunas pueden ser más eficaces que otras, por lo que, se propone una serie de enfoques de aspectos de calidad, como ser: consistencia, precisión, duplicidad, integridad, entre otros. Este trabajo es un inicio en la orientación a la calidad del dato público, por lo que se muestra la importancia de este para que los ciudadanos, aumenten el grado de confianza de la información brindada por el gobierno. El dato limpio y depurado, permite dar iniciación y apertura



a diversos procesos, como ser a la ingesta de datos, el cual trabaja con diferentes fuentes, estructuras o bien características para el correcto procesamiento de reutilización y aspectos varios en la interoperabilidad.

▪ **Publicación N°4:**

- **Fecha:** 19 al 30 de octubre de 2020.
- **Título:** “Análisis del procedimiento para la solicitud de información pública y tiempos de respuesta. Caso de Estudio: Ciudad Autónoma de Buenos Aires (Argentina)”.
- **Congreso:** 14° Simposio Argentino De Informática en el Estado - SIE que se desarrolló en el marco de las 49° Jornadas Argentinas De Informática (49° JAIIO). Jornadas Argentinas De Informática.
- **Lugar:** SADIO - Sociedad Argentina de Informática. Argentina. Virtual.
- **Autores:** Roxana Martínez, Rocío Rodríguez, Pablo Vera, Christian Parkinson.
- **Relación con la Tesis:** Este trabajo se enfoca en relevar la importancia de contar con datos abiertos que sean de calidad y permitan al ciudadano consultar información, tomar decisiones o incluso poder incorporarlos a herramientas propias para utilizarlos, lo que permitirá un rol más activo del ciudadano dándole transparencia a la gestión del organismo público. Para abordar el tema, se introduce en el artículo la diferencia entre datos públicos y abiertos, así como al marco regulatorio en Argentina. Se analiza un caso de estudio sobre el sistema de solicitud de información pública que actualmente tiene disponible desde la web el gobierno de la ciudad autónoma de Buenos Aires, con la finalidad de proponer mejoras al mismo. El funcionamiento de este sistema se puede evaluar por un lado considerando los datos requeridos y por otra parte analizando el porcentaje de solicitudes que fuera del plazo establecido por el organismo son atendidas o incluso quedan sin atenderse. Este análisis puede hacerse dado que se dispone de un dataset en el que quedan registradas las solicitudes de información (entre los años 2013 y 2019).

▪ **Publicación N°5:**

- **Fecha:** 05 y 06 de noviembre de 2020.
- **Título:** “Validación de la Calidad en Datos Abiertos con respecto a la detección de errores ortográficos utilizando la métrica del factor Syntactic Correctness”.
- **Congreso:** 8º Congreso Nacional de Ingeniería Informática – Sistemas de Información CoNalISI 2020.
- **Lugar:** Universidad Tecnológica Nacional, Facultad Regional San Francisco. Córdoba, Argentina. Virtual.
- **Autores:** Roxana Martínez; Christian Parkinson.
- **Relación con la Tesis:** En este trabajo se muestran distintos principios de calidad de los datos abiertos propuestos por organizaciones reconocidas a nivel internacional en la temática de gobierno abierto. Además, se proponen algunos aspectos a tener en cuenta en la calidad del dato a través de la métrica para la detección de errores ortográficos, utilizando el concepto del factor Syntactic Correctness. El objetivo y aporte de esta investigación fue incorporar y adaptar la herramienta más apropiada (según análisis comparativo realizado y verificado en este mismo estudio) con la inclusión de nuevas características y funciones de detección, en un desarrollo propio



de software, el cual realiza un control de la calidad de los datos abiertos desde el aspecto de la detección de errores ortográficos en los datos de cada uno de los campos del tipo cadena de texto que conforman un dataset analizado. Se tuvo en cuenta distintos criterios de análisis, que llevó al diseño de algoritmos en el prototipo para una mejor detección de falencias, como así de cantidad de ocurrencias por error. Otro aspecto en este trabajo fue el análisis de los campos completos, con el fin de verificar que todos fueron cargados, en caso contrario, se contabiliza como campo faltante. Este resultado, permite que un usuario utilice la herramienta de validación propuesta, y logre una mejor concientización de los aspectos que implican la calidad en los datos abiertos.

▪ **Publicación N°6:**

- **Fecha:** 05 y 06 de noviembre de 2020.
- **Título:** *“Propuestas de Buenas Prácticas para la Implementación de Smart City en el contexto de Datos Abiertos para la Salud Pública”.*
- **Congreso:** 8º Congreso Nacional de Ingeniería Informática – Sistemas de Información CoNalISI 2020.
- **Lugar:** Universidad Tecnológica Nacional, Facultad Regional San Francisco. Córdoba, Argentina. Virtual.
- **Autores:** Roxana Martínez; Martín Caruso; Diego López.
- **Relación con la Tesis:** En este trabajo contribuye en el relevamiento del contexto actual de los datos abiertos, datos públicos, transparencia, cuestiones organizativas gubernamentales en gobierno abierto y el marco normativo orientado al ambiente de la Salud Pública. Por otra parte, se incorporan aspectos de sobre ciudades inteligentes y se estudia la importancia de los datos públicos con respecto al tratamiento de estos entre organizaciones estatales y los ciudadanos. Se realiza un relevamiento de los trabajos más destacados en relación con los datos abiertos de salud pública y ciudades inteligentes en este mismo entorno, con el fin de establecer los criterios más importantes para lograr implementar una ciudad inteligente. Debido a esto, se indican una serie de propuestas en concepto de buenas prácticas para el tratamiento de datos abiertos, a nivel organizacional gubernamental y marco normativo en salud. Finalmente, se presentan recomendaciones de la Alianza para el Gobierno Abierto, Open Government Partnership (OGP) en lo que respecta a la salud pública para COVID-19 (enfermedad infecciosa causada por el coronavirus que se ha descubierto más recientemente).

▪ **Publicación N°7:**

- **Fecha:** 01 al 04 de diciembre de 2020.
- **Título:** *“Analysis of datasets and catalogs in government open portals of the Argentine Republic”.*
- **Congreso:** V Biennial Congress of IEEE Argentina Section.
- **Lugar:** IEEE. UTN Argentina, Resistencia. Virtual.
- **Autores:** Roxana Martínez; Rocío Rodríguez; Pablo Vera.
- **Relación con la Tesis:** En el presente trabajo se analiza la calidad de los datos en los datasets presentados en los sitios de portales abiertos de la República Argentina,



considerando al portal de la Ciudad Autónoma de Buenos Aires (C.A.B.A) y los organismos descentralizados del país. El objetivo es relevas que aspectos son los necesarios y mostrar los hallazgos frente al estudio de los diversos datasets publicados en los portales abiertos. Además, se muestran algunas consideraciones importantes a la hora de realizar el mantenimiento, gestión y la reutilización de éstos, con el fin de generar valor agregado a la comunidad. Finalmente, se relevan y validan los catálogos de datos abiertos, mediante una biblioteca de Python (pydatajson), para corroborar si cumplen con una estructura de catálogo estándar, propuesta por el gobierno de la República Argentina. Todo lo anteriormente dicho, conlleva a una mejor comprensión sobre qué elementos son esenciales en el tratamiento e identificación de los datos publicados en datasets.

▪ **Publicación N°8:**

- **Fecha:** 04 de abril de 2021.
- **Título:** *“Quality evaluation of government open data sets in Argentina using the HEVDA Validation Tool”*.
- **Journal:** Journal of Science and Research: Revista Ciencia e Investigación. ISSN 2528-8083, 6(2).
- **Autores:** Roxana Martínez; Claudia Pons; Rocío Rodríguez; Pablo Vera.

Relación con la Tesis: Este trabajo consiste en analizar la calidad de los conjuntos de datos disponibles en el portal web del sitio público oficial y nacional del país de Argentina. Esta investigación propone realizar un estudio de calidad aplicando la Herramienta de Validación de Calidad de Datos Abiertos (HEVDA), esta herramienta implementa métricas de calidad que miden el conjunto de datos seleccionado, lo que da como resultado un análisis de las fallas detectadas en el mismo; por ejemplo, permite detectar si existen errores, registros incompletos, tipos de redundancia, etc. Para explicar el encuadre de este trabajo se muestra un relevamiento de los aspectos que se involucran en este contexto: gobierno abierto, datos públicos abiertos, como así también la transparencia del gobierno. Por otro lado, muestra la importancia de mantener la calidad de los datos compartidos, ya que serán reutilizados en diferentes fuentes de datos y software, por lo que esta investigación se centra en los aspectos necesarios que favorecen la lectura y comprensión de los conjuntos de datos publicados en portales gubernamentales, lo que permite generar opinión pública y mostrar la trazabilidad de la gestión de los recursos gubernamentales.

<https://revistas.utb.edu.ec/index.php/sr/article/view/1179>

▪ **Publicación N°9:**

- **Fecha:** 13 de septiembre 2021.
- **Título:** *“Metrics proposal to measure the quality of governmental datasets”*.
- **Journal:** IEEE Latin America Transactions, 100(XXX).
- **Autores:** Roxana Martínez; Rocío Rodríguez; Pablo Vera.
- **Relación con la Tesis:** Se analizan los datos como un recurso indispensable para cualquier actividad de gestión pública, por lo que se indica la gran necesidad de mantenerlos actualizados y completos en los datasets que son disponibilizados por el



gobierno. Si bien se puede determinar que cada vez más organismos estatales se están embarcando en este nuevo concepto de gobierno abierto, aún queda un largo camino por recorrer. Este trabajo enfoca hacia la calidad de los datos abiertos de los portales de las entidades gubernamentales, y la necesidad de calidad en los datasets abiertos para que puedan ser mejor reutilizados, manipulados y distribuidos a los ciudadanos. Esta investigación se centra en la propuesta de métricas de calidad para el estudio de los contenidos de los conjuntos de datos publicados y así, obtener un análisis para mejorar su difusión e interoperabilidad entre bases de datos y software.
<https://latamt.ieeer9.org/index.php/transactions/article/view/5642>

▪ **Publicación N°10:**

- **Fecha:** 18 al 29 de octubre 2021.
- **Título:** *“Validación de Métricas propuestas de Calidad para el estudio de datos abiertos en base a criterios estándares: Aplicación HEVDA”.*
- **Congreso:** 50 JAIIO. Jornadas Argentinas de Informática.
- **Lugar:** Sociedad Argentina de Informática (SADIO). Virtual.
- **Autores:** Roxana Martínez; Claudia Pons; Rocío Rodríguez; Pablo Vera.
- **Relación con la Tesis:** 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.

▪ **Publicación N°11:**

- **Fecha:** 04 al 05 de noviembre de 2021.
- **Título:** *“Análisis de la Apertura de datos gubernamentales en los portales provinciales de la República Argentina: Aplicación HEVDA”.*
- **Congreso:** Congreso Nacional de Ingeniería Informática / Sistemas de Información (CoNaIISI). En modalidad virtual.
- **Lugar:** Universidad Tecnológica Nacional (UTN). Facultad Regional de Mendoza. Mendoza, Argentina.
- **Autores:** Roxana Martínez.
- **Relación con la Tesis:** Este trabajo presenta un análisis de los aspectos de calidad en los datasets publicados en los sitios web gubernamentales de las distintas provincias de la República Argentina, utilizando la herramienta propuesta desde un enfoque de análisis sobre los tipos de bloqueantes detectados en los conjuntos de datos públicos provinciales (Los tipos de bloqueantes se explicaron en la *sección 5.1.3. Aspectos Bloqueantes* de la presente tesis).



▪ **Publicación N°12:**

- **Fecha:** 15 de febrero 2022 – Invitación a presentación de extensión de Paper- Paper aceptado.
- **Título:** *“Detección de errores ortográficos para la validación de la calidad en datos abiertos gubernamentales para la métrica del factor syntactic correctness”.*
- **Estado:** En proceso de publicación.
- **Editorial:** Editora Artemis has open calls for the ebook “Ciências Socialmente Aplicáveis: Integrando Saberes e Abrindo Caminhos” (“Socially Applicable Sciences: Integrating Knowledge and Opening Paths”).
- **Lugar:** Curitiba, Brasil.
- **Autores:** Roxana Martínez.

▪ **Publicación N°13:**

- **Fecha:** 10 de marzo 2022 – Paper aceptado para publicación.
- **Título:** *“Quality Study of open government data related to COVID-19 in Latin America”.*
- **Estado:** En proceso de publicación.
- **Journal:** Journal of the School of Engineering of the Antioquia University, Colombia.
- **Lugar:** Universidad de Antioquia Medellín, Colombia.
- **Autores:** Roxana Martínez; Claudia Pons; Rocío Rodríguez; Pablo Vera.

▪ **Publicación N°14/Invitación:**

- **Fecha:** 15 de enero 2022 – A la fecha - Invited Article for extended version - conference paper.
- **Título:** En proceso de desarrollo.
- **Estado:** En proceso de publicación.
- **Journal indexada en Scopus:** Special Issue on Innovation in Computing, Engineering Science & Technology organized by Advances in Science, Technology and Engineering Systems Journal (ASTESJ).
- **Lugar:** USA.
- **Autores:** Roxana Martínez; Claudia Pons.



Anexos

Anexo A: ISO/IEC 25012

En esta sección se muestra el detalle de la calidad de datos de la Norma ISO/IEC 25012 extraída de [ISO12].

Calidad de Datos Inherente

Exactitud: 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:

Exactitud Sintáctica: cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado sintácticamente correcto.

Exactitud Semántica: cercanía de los valores de los datos a un conjunto de valores definidos en un dominio considerado semánticamente correcto.

Compleitud: 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.

Consistencia: 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.

Credibilidad: Grado en el que los datos tienen atributos que se consideran ciertos y creíbles en un contexto de uso específico. La credibilidad incluye el concepto de autenticidad (la veracidad de los orígenes de datos, atribuciones, compromisos).

Actualidad: Grado en el que los datos tienen atributos que tienen la edad correcta en un contexto de uso específico.

Calidad de Datos Inherente y Dependiente del Sistema

Accesibilidad: Grado en el que los datos pueden ser accedidos en un contexto específico, particularmente por personas que necesiten tecnologías de apoyo o una configuración especial por algún tipo de discapacidad.

Conformidad: Grado en el que los datos tienen atributos que se adhieren a estándares, convenciones o normativas vigentes y reglas similares referentes a la calidad de datos en un contexto de uso específico.

Confidencialidad: Grado en el que los datos tienen atributos que aseguran que los datos son sólo accedidos e interpretados por usuarios autorizados en un contexto de uso específico.



Eficiencia: Grado en el que los datos tienen atributos que pueden ser procesados y proporcionados con los niveles de rendimiento esperados mediante el uso de cantidades y tipos adecuados de recursos en un contexto de uso específico.

Precisión: Grado en el que los datos tienen atributos que son exactos o proporcionan discernimiento en un contexto de uso específico.

Trazabilidad: Grado en el que los datos tienen atributos que proporcionan un camino de acceso auditado a los datos o cualquier otro cambio realizado sobre los datos en un contexto de uso específico.

Comprensibilidad: Grado en el que los datos tienen atributos que permiten ser leídos e interpretados por los usuarios y son expresados utilizando lenguajes, símbolos y unidades apropiados en un contexto de uso específico. Cierta información sobre la comprensibilidad puede ser expresada mediante metadatos.

Calidad de Datos Dependiente del Sistema

Disponibilidad: Grado en el que los datos tienen atributos que permiten ser obtenidos por usuarios y/o aplicaciones autorizadas en un contexto de uso específico.

Portabilidad: Grado en el que los datos tienen atributos que les permiten ser instalados, reemplazados o eliminados de un sistema a otro, preservando el nivel de calidad en un contexto de uso específico.

Recuperabilidad: Grado en el que los datos tienen atributos que permiten mantener y preservar un nivel específico de operaciones y calidad, incluso en caso de fallos, en un contexto de uso específico.



Anexo B: Principios Internacionales de los datos abiertos

En esta sección se muestran los principios establecidos por:

a) Open Data Charter, Carta Internacional de Datos Abiertos

Los seis principios del Open Data Charter [ODC21], serán la base para el acceso a los datos y para la divulgación y uso de éstos:

1. Abrir por defecto: Quiere decir que los datos están abiertos a todos los ciudadanos. Los gobiernos deben justificar los datos que se mantienen cerrados, por ejemplo, por razones de seguridad o protección de datos.

2. Oportuno y completo: Los datos abiertos sólo son valiosos si siguen siendo relevantes. Por otro lado, la información se publica de forma rápida y completa, lo que es fundamental para su éxito. En la medida de lo posible, los gobiernos deberían proporcionar datos en su forma original y sin modificaciones.

3. Accesible y utilizable: Garantizar que los datos sean legibles por máquina y fáciles de encontrar. Los portales son una forma de lograrlo. Pero también es importante pensar en la experiencia del usuario de quienes acceden a los datos, incluidos los formatos de archivo en los que se proporciona la información.

4. Comparable e interoperable: Cuantos más conjuntos de datos de calidad tenga acceso y más fácil sea para ellos hablar entre ellos, más valor potencial podrá obtener de ellos. Los estándares de datos cumplen un rol fundamental.

5. Mejorar la gobernanza y la participación ciudadana: Los datos abiertos tienen la capacidad de permitir que los ciudadanos (y otros en el gobierno) tengan una mejor idea de lo que están haciendo los funcionarios y los políticos. Esta transparencia puede mejorar los servicios públicos y ayudar a que los gobiernos rindan cuentas.

6. Para el desarrollo y la innovación inclusivos: los datos abiertos pueden ayudar a impulsar el desarrollo económico inclusivo. Por ejemplo, un mayor acceso a los datos puede hacer que la agricultura sea más eficiente o puede usarse para abordar el cambio climático.

b) Open Government Data, Datos de Gobierno Abierto

La Open Government Data, Datos de Gobierno Abierto [OGD21], tiene como misión brindar principios para que los datos gubernamentales, se identifiquen como datos abiertos. Los datos gubernamentales se considerarán abiertos si se hacen públicos de una manera que cumpla con los principios siguientes:

1. Completa: Todos los datos públicos están disponibles. Los datos públicos son datos que no están sujetos a limitaciones válidas de privacidad, seguridad o privilegios. Los ciudadanos tienen derecho a conocer y reutilizar el contenido de los datos que contienen los gobiernos.

2. Primaria: Los datos se recopilan en la fuente, con el mayor nivel posible de granularidad, no en forma agregada o modificada.

3. Oportuna: Los datos están disponibles tan rápido como sea necesario para preservar el valor de los datos.



4. **Accesible:** Los datos están disponibles para la más amplia gama de usuarios para la más amplia gama de propósitos.
5. **Procesable a máquina:** Los datos están razonablemente estructurados para permitir un procesamiento automatizado.
6. **No discriminatorio:** Los datos están disponibles para cualquier persona, sin necesidad de registrarse.
7. **No propietario:** Los datos están disponibles en un formato sobre el que ninguna entidad tiene control exclusivo.
8. **Sin licencia:** Los datos no están sujetos a ningún derecho de autor, patente, marca registrada o regulación de secreto comercial. Se pueden permitir restricciones razonables de privacidad, seguridad y privilegios.



Anexo C: Funciones secundarias de la Herramienta HEVDA

En esta sección se muestran las funciones adicionales que fueron desarrolladas en la aplicación a modo de resumen.

Al validar un datasets desde la herramienta HEVDA, se puede obtener información complementaria sobre la estructura del dataset:

- Nombre del dataset procesado;
- Tiempo/Duración del proceso;
- Cantidad de Registros encontrados;
- Cantidad de columnas encontradas;
- Validación si el archivo cumple con el tipo de formato de archivo CSV;
- Validación si el archivo cumple con el criterio de separador de campos por "," (coma);
- Validación si el primer registro del archivo contiene los títulos de las columnas;
- Validación si el dataset tiene nombres de columnas repetidas;
- Validación si el dataset contiene columnas afectadas con elementos que inician o finalizan con espacios en blanco;
- Validación si el archivo presenta codificación UTF-8;

En la Figura Anexo C.1, se muestra una captura de pantalla con un análisis estructural a modo de resumen del dataset analizado, en este caso para el dataset "produccion-de-bioetanol-por-insumo-.csv". Además, se muestra un símbolo de cruz con color rojo, que indica que existen columnas con elementos que inician o finalizan en blanco, para el resto de las validaciones se tiene una tilde en color verde indicando que para cada una de éstas son correctas.



Figura Anexo C.1 – Pantalla de resumen de la Herramienta HEVDA.



Como última funcionalidad, se muestra un resumen de las métricas críticas que cumplen o no cumplen a modo de resumen visual, tal como se observa en la Figura Anexo C.2.

Datos Abiertos

Métricas Críticas ▾

Métricas No Críticas ▾

RESUMEN DE LA VERIFICACIÓN DE CALIDAD DE DATOS

Para más detalle, realizar clic en cada opción de métrica del menú

ANÁLISIS ESTRUCTURAL DEL DATASET

Dataset procesado: produccion-de-bioetanol-por-insumo-.csv
Tiempo/Duración del Proceso: 00:00:00.2343722
Cantidad de Registros: 29 Cantidad de columnas: 9

- ✓ El archivo cumple con el tipo de formato de archivo CSV
- ✓ El archivo CSV cumple con el criterio de separador de campos por ","
- ✓ El primer registro del archivo contiene los títulos de las columnas
- ✓ El dataset no tiene nombres de columnas repetidos
- ✗ El dataset contiene 1 columnas afectadas con elementos que inician o finalizan con espacios en blanco
- ✓ Codificación del archivo: UTF-8

RESUMEN DE MÉTRICAS CRÍTICAS

- ✓ Validación de Formato de Datos Decimal
- ✓ Validación de Registros Duplicados
- ✓ Validación de Detección de Datos Incompletos
- ✗ Validación de Catacteres Inválidos

Figura Anexo C.2 – Pantalla de resumen de métrica críticas en HEVDA.



Anexo D: Portales de datos abiertos gubernamentales

En esta sección se muestran las capturas/evidencias de pantalla de los portales de datos abiertos gubernamentales tratados en la recolección de datos del capítulo 5 al 13 de enero del 2021.

En la Figura Anexo D.1, se muestra el portal de datos públicos del Ministerio de Salud de Argentina [SAL21] con un total de 41 datasets.



Figura Anexo D.3 – Portal del Ministerio de Salud de Argentina [SAL21].

En la Figura Anexo D.2, se muestra el portal de datos públicos de Diputados de la Argentina [DIP21] con un total de 29 datasets.

En la Figura Anexo D.3, se muestra el portal de datos públicos del Ministerio de Justicia y Derechos Humanos de la Argentina [MIN21] con un total de 61 datasets.

En la Figura Anexo D.4, se muestra el portal de datos públicos del Portal de la Ciudad de Buenos Aires [BAC21] con un total de 398 datasets.

En la Figura Anexo D.5, se muestra el portal nacional de datos públicos de Argentina [AUD21] con un total de 1013 datasets.



Figura Anexo D.4 – Portal de Diputados de la Argentina [DIP21].

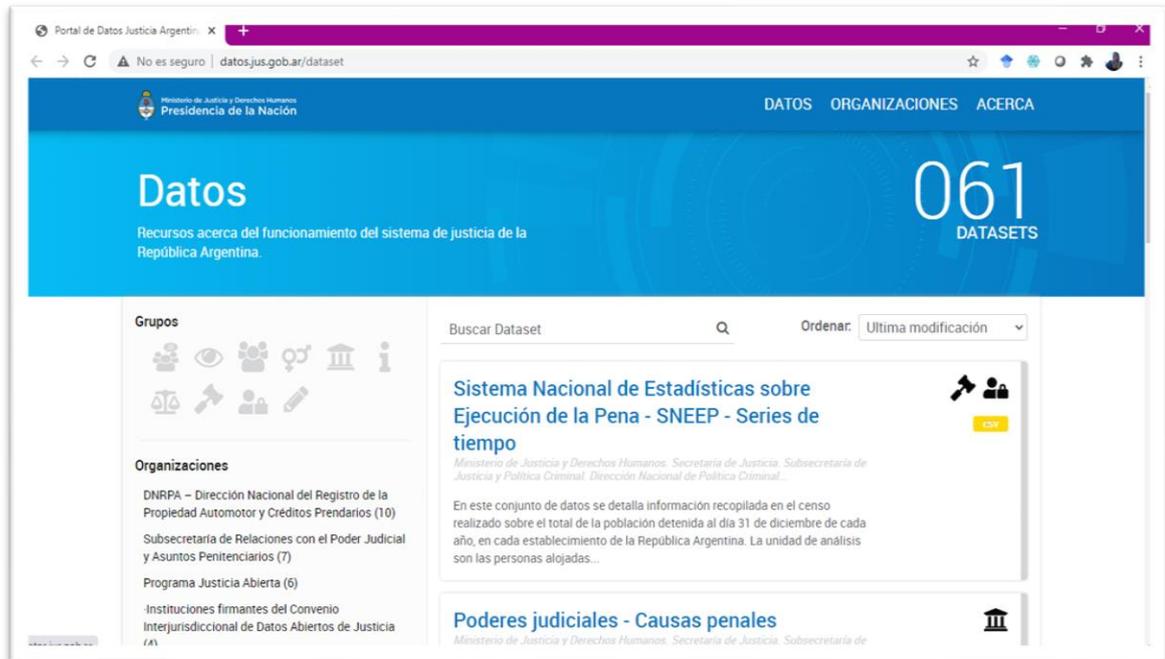


Figura Anexo D.5 – Portal del Ministerio de Justicia y Derechos Humanos de Argentina [MIN21].

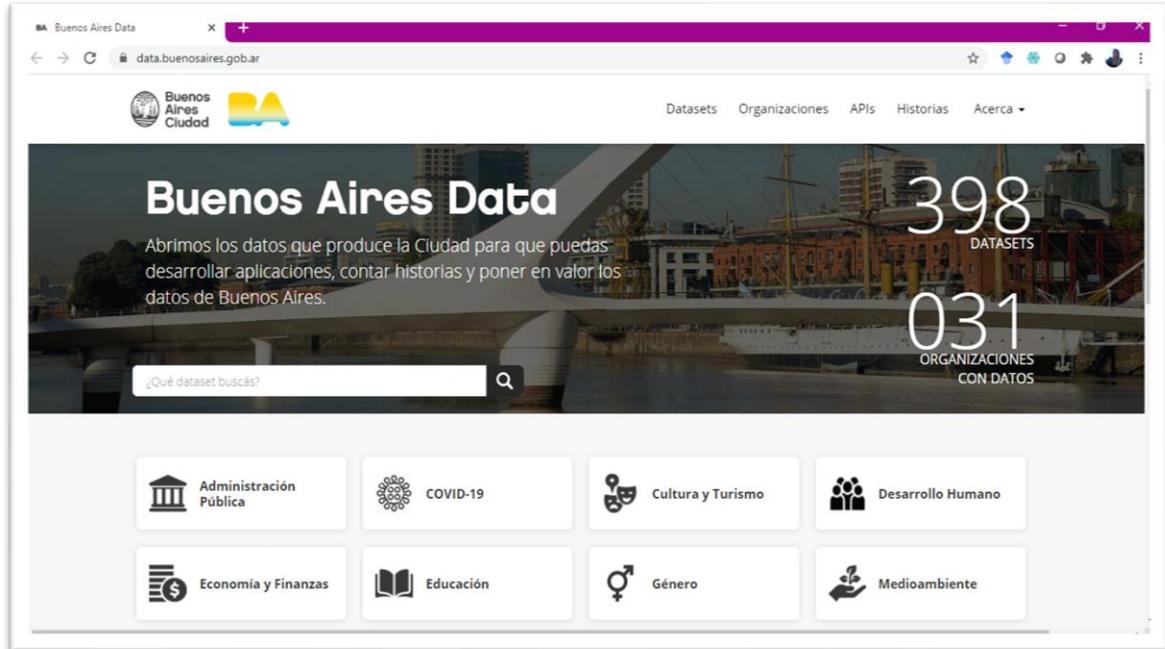


Figura Anexo D.6 – Portal de la Ciudad de Buenos Aires [BAC21].

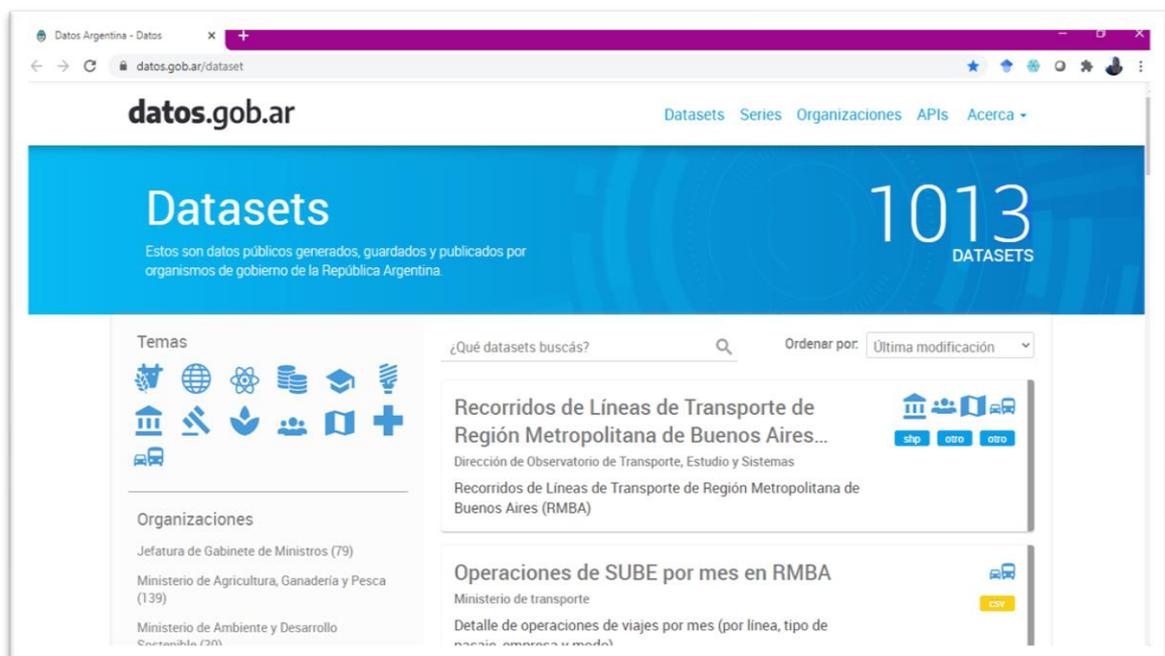


Figura Anexo D.7 – Portal Nacional de Argentina [AUD21]



Anexo E: Detalle de los resultados arrojados con HEVDA

En esta sección se muestra el detalle de los resultados obtenidos con la validación de la herramienta HEVDA, identificando los portales de datos abiertos gubernamentales, el nombre del dataset, si es dataset es del tipo bloqueante y en caso de que no sea bloqueante, si cumple o no con las métricas 1 a la 8 inclusive, siendo:

- METRICA 1: NUMÉRICOS DECIMALES;
- METRICA 2: REGISTROS DUPLICADOS;
- METRICA 3: DATOS INCOMPLETOS;
- METRICA 4: CARACTERES INVÁLIDOS;
- METRICA 5: REDUNDANCIA EN EL DOMINIO DE VALORES DE UNA COLUMNA;
- METRICA 6: REDUNDANCIA ENTRE CAMPOS DE UNA MISMA FILA;
- METRICA 7: DETECCIÓN DE VALORES ID;
- METRICA 8: CAMPOS TRIVIALES;

Y los Tipos de Bloqueantes, siendo:

- Tipo 1 - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA);
- Tipo 2 - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS;
- Tipo 3 - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA);
- Tipo 4 - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET;
- Tipo 5 - EL ARCHIVO POSEE NOMBRES DE TÍTULOS REPETIDOS;

En la Tabla Anexo E.1, se muestran los datasets utilizados para el portal de datos Abiertos del Ministerio de Salud [SAL21].

Tabla Anexo E.1 – Detalle de datasets validados del Ministerio de Salud.

NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
1	<i>expresiones-voluntad-provincia-20180725.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
2	<i>Covid19Determinaciones.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
3	<i>beneficiarios-inscriptos-programa-sumar-con-cobertura-efectiva-basica-20181218.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
4	<i>legisalud2018-2.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
5	<i>establecimientos-062018.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
6	<i>dengue-zika-nacional-201807.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE					
7	<i>recetas-medicamentos-esenciales.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
8	<i>contrataciones-compras-licitaciones-20180921.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
9	<i>certificacion-especialistas-provincia-2016.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
10	<i>tasa-mortalidad-infantil.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE

En la Tabla Anexo E.2, se muestran los datasets utilizados para el portal de datos Abiertos de Diputados de la Nación Argentina [DIP21].

Tabla Anexo E.2 – Detalle de datasets validados de Diputados de la Nación Argentina.

NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
11	<i>desarraigo1.4.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
12	<i>ejecucion-al-31-de-diciembre2019.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
13	<i>remuneraciones1.4.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
14	<i>sesiones1.4.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
15	<i>sumario-leyes1.3.csv</i>	SI - EL ARCHIVO POSEE DOBLE								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		CARACTER " (COMILLA)								
16	resultados-proyectos1.4.csv	NO	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
17	giros-comisiones1.5.csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE

En la Tabla Anexo E.3, se muestran los datasets utilizados para el portal de datos Abiertos de Justicia Argentina [MIN21].

Tabla Anexo E.3 – Detalle de datasets validados de Justicia Argentina.

NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
18	base-infoleg-normativa-nacional-muestreo.csv	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
19	bienes-secuestrados-decomisados-durante-proceso-penal-20201231.csv	NO	CUMPLE	NO CUMPLE						
20	codificacion-delitos-codigo-penal-argentino-20191011.csv	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
21	cv-magistrados-poder-judicial-ministerio-publico-20210112.csv	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
22	indice-archivos-recibidos-poder-judicial-no-penal-2020.csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
23	indice-archivos-recibidos-poder-judicial-penal-2020.csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
24	listado-comunidades-indigenas-20201120.csv	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
25	listado-mediadores-202012.csv	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
26	llamados-atendidos-abuso-sexual-202012.csv	NO	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
27	llamados-atendidos-violencia-familiar-202012.csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
28	registro-de-femicidios-20200109.csv	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
29	registro-viajes-financiados-terceros-20201218.csv	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
30	sneep-poblacion-total.csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
31	solicitudes-acceso-informacion-publica-20210107.csv	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
32	ternas-seleccion-magistrados-poder-judicial-ministerio-publico-20210112.csv	NO	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE				

En la Tabla Anexo E.4, se muestran los datasets utilizados para el portal de datos Abiertos de la Ciudad de Buenos Aires, Argentina [BAC21].

Tabla Anexo E.4 – Detalle de datasets validados de la Ciudad de Buenos Aires, Argentina.

NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
33	solicitudes-y-reclamos-de-defensoria-del-pueblo.csv	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
34	resultados-elecciones-legisladores-generales-2017.csv	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
35	denuncias-defensa-del-consumidor.csv	SI - EL ARCHIVO NO CUMPLE								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		CON EL FORMATO CSV DE SEPARADOR (COMA)								
36	<i>ciudad-limpia-campana-responsable.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
37	<i>embajadas.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
38	<i>info-rrhh-por-genero.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
39	<i>llamadas_linea_144.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
40	<i>casos_penales_contravencion_industria_y_comercio_genero.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
41	<i>contactos_botiquin_covid_19.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
42	<i>dataset_flujo_vehicular.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
43	<i>dataset_reporte_covid_sitio_gobierno.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
44	<i>dataset_traslados_covid_19.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
45	<i>dataset_viajes_sube.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
46	<i>distanciamiento_covid19.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
47	<i>ubicacion-ufus.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
48	<i>llamados_107_covid.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
49	<i>baset-2019.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
50	<i>encuesta-joven-2014-acceso-a-tecnologias-y-usos-digitales.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
51	<i>espacios-culturales.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
52	<i>lugares-de-culto.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
53	<i>programacion.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
54	<i>programacion-actual-html.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
55	<i>sedes-vicarias.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
56	<i>visitas-guiadas.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
57	<i>barrios_populares_badata_WGS84.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE				
58	<i>cond_pob_hog_anno_s_jefe_ringr_ctotcaba_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
59	<i>espacios-verdes-porhabitante-porcomuna.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
60	<i>informacion-censal-por-radio-2001.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
61	<i>porc_jeffas_hog_anno_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
62	<i>porc_lic_var_anno_g_edad_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
63	<i>prom_t_simul_merc_16sexo_anno_g_edad_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
64	<i>unidades-territoriales-deinclusion-urbana.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
65	<i>bre_ing_prom_anno_calif_ocup_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
66	<i>compras-coronavirus.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
67	<i>concesiones.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
68	<i>empresas-de-desinfeccion-y-desinfestacion.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
69	<i>empresas-habilitadas.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
70	<i>jer_ocup_epfsexo_ _annio_jer_ocup_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
71	<i>recaudacion-impositiva.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
72	<i>registro-administradores-de-consorcio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
73	<i>cond_asis_esc_edad _sexo_ annio_ c_ a sist_limpio.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
74	<i>datos_esi_ogp_2018 .csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
75	<i>distritos-escolares.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
76	<i>esp_vid_escsexo_ annio_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
77	<i>inscripciones-escolares-2015.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		SEPARADOR (COMA)								
78	<i>interpretes-de-lenguaje-de-senas.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
79	<i>programa-de-educacion-olimpica.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
80	<i>universidades.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
81	<i>est_cond_asis_edad_anno_sex_asis_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
82	<i>med_ing_hora_ocup_sex_anno_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
83	<i>prom_t_simul_cuid_16_sex_anno_g_edad_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
84	<i>tas_femic_anno_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
85	<i>tas_glob_fec_anno_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
86	<i>tas_part_lad_sex_anno_g_g_activ_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
87	<i>tip_viol_pare_anno_g_edad_t_viol_obs_limpio.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
88	<i>visxvec_est_sal_sex_anno_residen_limpio.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
89	<i>calidad-aire.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
90	<i>consultoras.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
91	<i>informacion-meteorologica-2011.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
92	<i>generadores.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
93	<i>mediciones-rn-puntos-de-medicion.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
94	<i>rutas_recoleccion_residuos_humedos.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
95	<i>transformadores-pcb.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
96	<i>zonas_de_recoleccion_de_residuos_limpios.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
97	<i>acompanantes-autorizados-de-servicios-escolares.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
98	<i>estaciones-accesibles.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
99	<i>estaciones-de-ferrocarril.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
100	<i>estado-de-flota.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
101	<i>frecuencia_subte.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
102	<i>agencias-de-radio-taxi.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
103	<i>garajes-comerciales.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
104	<i>registro-historico-del-precio-del-boleto.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
105	<i>centros_de_salud_nivel_1_BADATA_WG_S84.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
106	<i>centros-de-atencion-veterinaria.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
107	<i>donacion.csv</i>	SI - EL ARCHIVO NO								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
108	<i>establecimientos-elecciones-2017.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
109	<i>estaciones-saludables.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
110	<i>farmacias.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
111	<i>geriatricos.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
112	<i>hospitales.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
113	<i>porc_sida_muj_anno_g_edad_limpio.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
114	<i>concluidas17.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
115	<i>departamentos-comisarias-comunales.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
116	<i>fideicomisos.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
117	<i>ingresadas17.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
118	<i>personas-buscadas.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
119	<i>registro-de-fabricantes-reparadores-y-recargadores-de-extintores.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
120	<i>senderos-escolares.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
121	<i>tobilleras_limpio.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE				
122	<i>vigiladores-de-locales-bailables-historico.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
123	<i>rivolta.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
124	<i>sumideros-zona-1.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
125	<i>estadios.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
126	<i>federaciones.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
127	<i>iglesias.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
128	<i>lib.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
129	<i>mapa-manzanas.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
130	<i>precio-de-terrenos-2001.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
131	<i>restricciones.csv</i>	NO	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE

En la Tabla Anexo E.5, se muestran los datasets utilizados para el portal de datos Abiertos de la República Argentina [AUD21].

Tabla Anexo E.5 – Detalle de datasets validados de la República Argentina.

NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
132	<i>huertas-familiares-por-provincia.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
133	<i>cna-02-residentes-relacion-productor-provincia-total.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
134	<i>produccion_leche_mes.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
135	<i>faena_aviar.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
136	<i>serie-tiempo-consumo_carne_aviar.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
137	<i>ingreso-de-trigo-candeal-con-destino-a-la-industria-balanceadora-.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
138	<i>superficie-nacional-fiscalizada-especies-2017-2018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
139	<i>ingreso-de-trigo-pan-con-destino-a-la-industria-balanceadora-.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
140	<i>lacteos-exportaciones-anual-dolares-1989-2009.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
141	<i>indice-de-costo-de-pollos-parrilleros.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
142	<i>movimiento-ciervos-provincia-2013-2018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
143	<i>productos-lacteos-liquidos-1989-2006.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
144	<i>extracciones-forestales-1998-2017.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
145	<i>importaciones-de-azucar.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		UNO DE SUS REGISTROS								
146	<i>existencias-porcinas-provincia-departamento-2008-2019.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
147	<i>existencias-bovinas-provincia-departamento-2008-2019.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
148	<i>expo-2020-vegetales-certificados-senasa.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
149	<i>produccion-de-harina-de-trigo-pan-.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
150	<i>ingreso-de-trigo-pan-con-destino-a-la-industria-molinera-.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
151	<i>produccion-de-pellets-de-trigo-pan-.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
152	<i>registro-nacional-cultos.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		COLUMNAS EN CADA UNO DE SUS REGISTROS								
153	<i>registro-institutos-vida-consagrada.csv</i>	NO	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
154	<i>lista-sedes.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
155	<i>solicitudes-informacion-publica-recibidas-2017-acumar.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
156	<i>residencias-otorgadas---2-trimestre---ano-2020.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
157	<i>Sancionesaoperador esporreclamosinicia dosporlosusuarios_2651791611022230446.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
158	<i>ConectividadalserviciodelInternet_2776151611021035282.csv</i>	NO	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE				
159	<i>Fomecamontosaprobadosypagadosporlineadeconcurso_2692491611022457619.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
160	<i>Reclamos por operador y motivo de reclamo.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
161	<i>Registro de prestadores de servicios postales.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
162	<i>Denuncias_reclamos (Denuncias mensuales recibidas).csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
163	<i>Registro de Proveedores Postales.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		DE LAS COLUMNAS DEL DATASET								
164	rep_autorizadas.csv.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
165	Television.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
166	Penetración del Internet fijo por provincia (accesos por cada 100 habitantes).csv	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
167	Reclamos de usuarios de servicios postales por operador.csv	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
168	Total nacional de accesos a Internet fijo por tipo de tecnología.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
169	Total nacional de accesos a telefonía fija.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
170	Unidades mensuales vendidas de servicios postales del correo oficial y privados.csv	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
171	Telefonía móvil (Total nacional de accesos a telefonía móvil).csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
172	Telefonía móvil (Telefonía móvil minutos salientes (miles)).csv	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
173	servicios_postales.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
174	Total nacional de accesos a Internet fijo por banda ancha y banda angosta.csv	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
175	Penetración total nacional de la televisión por	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
	<i>suscripción y satelital.csv</i>		NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
176	<i>Denuncias_reclamos.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
177	<i>Telefonia_movil.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
178	<i>Denuncias_reclamos (Reclamos mensuales recibidos por motivo de reclamo).csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
179	<i>Internet_Ingresos.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
180	<i>Penetración nacional de la telefonía móvil (accesos por cada 100 habitantes).csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
181	<i>Velocidad promedio de bajada de Internet (nacional).csv</i>	SI - EL ARCHIVO POSEE NOMBRES DE TÍTULOS REPETIDOS								
182	<i>Internet_Penetración.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
183	<i>Penetración total nacional de la telefonía fija por trimestre.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
184	<i>Personal ocupado por trimestre en el correo oficial y en operadores privados.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
185	<i>Número y monto de sanciones a operadores por reclamos de usuarios por motivo.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
186	<i>autorizaciones.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
187	<i>pagos_fomeca_v2.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
188	<i>graph_licencias_audiov.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
189	<i>Cantidad autorizaciones radioeléctricas.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
190	<i>algodonera-textil.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
191	<i>ejecucion-fisica-trimestral_2017_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
192	<i>ejecucion-fisica-trimestral_2016_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
193	<i>ejecucion-fisica-trimestral_2015_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
194	<i>ejecucion-fisica-trimestral_2014_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
195	<i>ejecucion-fisica-trimestral_2013_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
196	<i>ejecucion-fisica-trimestral_2012_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
197	<i>ejecucion-fisica-trimestral_2011_1.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
198	<i>ejecucion-fisica-trimestral_2010_1.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE					
199	<i>mapeo_tematica_gero_fin1TRIM20.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
200	<i>mapeo_tematica_gero_fin1SEM.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
201	<i>pc-proy18-finfun.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
202	<i>pc-proy21-finfun.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
203	<i>relevamiento-expectativas-mercado-rem.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
204	<i>indice-precios-al-consumidor-apertura-por-capitulos-base-diciembre-2016-mensual.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
205	<i>ipc-categorias-tasas-variacionmensual-basedic-2016.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
206	<i>producto-interno-bruto-precios-corrientes-valores-anuales-base-2004.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
207	<i>pib-agricultura-caza-silvicultura-pesca-australes-precios-1970-trimestral.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
208	<i>indice-precios-al-consumidor-nivel-general-base-diciembre-2016-trimestral.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
209	<i>activos-externos-componentes-datos-historicos.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
210	<i>argentina-trabajadores-activos-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
211	<i>hacemos-futuro-trabajadores-activos-2020-02-29.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
212	<i>ellas-hacen-titulares-activos-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
213	<i>centrodesarrolloinfantil-2021-01-01.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
214	<i>argentina-trabajatitulares-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
215	<i>hacemos-futurotitulares-2020-02-29.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
216	<i>piddef.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
217	<i>potenciar-trabajototal-titulares-2020-12-31.csv</i>	NO	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE				
218	<i>argentina-trabajatotal-titulares-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
219	<i>ellas-hacen-titulares-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
220	<i>hacemos-futurototal-titulares-2020-02-29.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
221	<i>montos-mensuales-programas-transferencia-directa-2018-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
222	<i>canales-de-tv-abierta.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
223	<i>paeetapa2020-09-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
224	<i>paeetapa12020-09-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
225	<i>participacion-de-puestos-de-trabajo-registrados-y-no-registrados-y-no-asalariados-por-sector-de-ac.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
226	<i>montos-mensuales-programas-transferencia-directa-2020-02-29.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
227	<i>participacion-de-puestos-de-trabajo-por-sector-cultural-en-el-total-de-cultura..csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
228	<i>participacion-de-puestos-de-trabajo-culturales-en-el-total-de-la-economia-por-tipo-de-trabajo.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
229	<i>montos-mensuales-programas-transferencia-directa-2021-01-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
230	<i>ventas-de-productores-de-glp.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
231	<i>capitulo-iv-pozos.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
232	<i>pronstico-de-produccin-de-gas-natural-anterior-al-2009.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
233	<i>produccin-de-glp-condensado-y-gasolina-</i>	SI - EL ARCHIVO NO CUMPLE								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
	<i>estabilizada-produccion-de-glp.csv</i>	CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
234	<i>pronstico-de-produccion-de-petrleo-anterior-al-2009.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
235	<i>produccion-de-glp-por-planta-ant-2009.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
236	<i>datos-de-fractura-de-pozos-de-hidrocarburos-adjunto-iv-actualizacion-diaria.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE				
237	<i>compras-de-distribuidor-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
238	<i>ventas-de-petroquimicas-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
239	<i>compras-de-aerosoleras-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
240	<i>compras-de-petroquimicas-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
241	<i>series-ben-1960-2016.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
242	<i>ventas-de-comercializadores-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
243	<i>compras-de-comercializador-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
244	<i>compras-de-productores-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
245	<i>generacion-elctrica-centrales-de-generacion.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
246	<i>compras-de-consumidores-de-glp.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
247	<i>Precios de Gas Natural - Res 1 2018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
248	<i>biocombustible-precios-de-bioetanol.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
249	<i>categoras-de-empresas.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
250	<i>pauta-oficial-2014.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
251	<i>pauta-oficial-2015.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
252	<i>pauta-oficial-primero-primero-2017.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
253	<i>pauta-oficial-segundo-segundo-2017.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
254	<i>pauta-oficial-segundo-segundo-2016.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
255	<i>pauta-oficial-primero-primero-2019.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
256	<i>pauta-oficial-segundo-segundo-2018.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
257	<i>pauta-oficial-primero-primero-2018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
258	<i>pauta-oficial-primero-primero-2016.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
259	<i>pauta-oficial-segundo-segundo-2019.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
260	<i>dataset_mop.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
261	<i>publicidad-oficial-2020.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
262	<i>salarios-2019.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
263	<i>salarios-2018.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
264	<i>dataset_hospitales.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
265	<i>salarios-2017.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
266	<i>salarios-2016.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
267	<i>salarios-2020.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
268	<i>dataset_diccionario.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
269	<i>internos-spf-condenados-202012.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
270	<i>ternas-seleccion-magistrados-poder-judicial-ministerio-publico-20210112.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
271	<i>cv-magistrados-poder-judicial-ministerio-publico-20210112.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
272	<i>internos-spf-procesados-202012.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
273	<i>solicitudes-condicion-legitimo-usuario-armas-fuego-202101-quincena-1.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
274	<i>llamados-atendidos-abuso-sexual-202012.csv</i>	NO	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
275	<i>llamados-atendidos-violencia-familiar-202012.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
276	<i>poderes-judiciales-causas-penales-muestreo.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
277	<i>solicitudes-autorizacion-portacion-armas-fuego-202101-quincena-1.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
278	<i>intervenciones-domiciliarias-abuso-sexual-202012.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
279	<i>sneep-poblacion-total.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
280	<i>sneep-genero-varones.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
281	<i>sneep-situacion-legal-condenados-condenadas.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
282	<i>sneep-situacion-legal-procesados-procesadas.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
283	<i>sneep-edad-18-a-20.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
284	<i>sneep-genero-mujeres.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
285	<i>sneep-situacion-legal-otra-situacion.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
286	<i>sneep-edad-menores-18.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
287	<i>internos-spf-inimputables-202012.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
288	<i>sneep-genero-trans.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
289	<i>tabla-ia.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
290	<i>superficie-incendiada-provincias-tipo-de-vegetacion.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
291	<i>incendios-cantidad-causas-provincia.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
292	<i>muestreo-calidad-de-agua-riiglo-primavera-2019.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		FORMATO CSV DE SEPARADOR (COMA)								
293	<i>muestreo-calidad-de-agua-riiglo-verano-2019.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
294	<i>muestreo-calidad-de-agua-riiglo-otono-2019.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
295	<i>extracciones_productos_forestales-parques-nacionales-m3.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
296	<i>muestreo-calidad-de-agua-riiglo-invierno-2014.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
297	<i>muestreo-calidad-de-agua-riiglo-invierno-2013.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
298	<i>muestreo-calidad-de-agua-riiglo-primavera-2013.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
299	<i>muestreo-calidad-de-agua-riiglo-primavera-2014.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
300	<i>muestreo-calidad-de-agua-riiglo-otono-2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
301	<i>muestreo-calidad-de-agua-riiglo-invierno-2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
302	<i>muestreo-calidad-de-agua-riiglo-verano-2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
303	<i>muestreo-calidad-de-agua-riiglo-primavera-2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
304	<i>superficie-incendiada-parques-nacionales-tipo-de-vegetacion.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
305	<i>materiales-educacion-ambiental.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		CSV DE SEPARADOR (COMA)								
306	<i>incendios-cantidad-causas-parques-nacionales.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
307	<i>materiales-comunicacion-ambiental.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
308	<i>nombres-1935-1939.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
309	<i>nombres-1930-1934.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
310	<i>renabap-2020-11-20.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE				
311	<i>nombres-1925-1929.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
312	<i>datos-de-produccion-vegetal-de-los-naf.csv</i>	SI - EL ARCHIVO NO POSEE UNA PRIMERA FILA DE TÍTULOS/NOMBRES DE LAS COLUMNAS DEL DATASET								
313	<i>indice-calidad-vida.csv</i>	SI - EL ARCHIVO POSEE DOBLE CARACTER " (COMILLA)								
314	<i>nombres-1920-1924.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
315	<i>dataset_inventario_agryout_1990_2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
		FORMATO CSV DE SEPARADOR (COMA)								
316	<i>actividades-del-nucleo-de-agricultura-familiar-por-grupos-de-produccion.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
317	<i>integrantes-del-naf-segun-el-nivel-educativo-alcanzado.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
318	<i>indice-educacion.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
319	<i>integrantes-del-naf-por-composicion-individual.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
320	<i>indice-vivienda.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
321	<i>indice-entorno.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
322	<i>padron-de-operadores-organicos-certificados.csv</i>	NO	CUMPLE	NO CUMPLE						
323	<i>indice-salud.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
324	<i>producciones-de-los-nucleos-de-agricultura-familiar-por-tipo-de-animales.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
325	<i>regulacion_provincial_fitosanitarios.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
326	<i>dataset_inventario_agryout_1990_2016_definiciones.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
327	<i>establecimientos-062018.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
328	<i>asentamientos.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE				
329	<i>localidades.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
330	<i>localidades-censales.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
331	<i>onc_obras_decreto_1169_18.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
332	<i>municipios.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
333	<i>ign_municipio.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
334	<i>departamentos.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
335	<i>ign_departamento.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
336	<i>ign-pais.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
337	<i>indicadores-georef.csv</i>	NO	CUMPLE							
338	<i>centrosdereferencia-2021-01-01.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
339	<i>provincias.csv</i>	NO	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
340	<i>ign_provincia.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
341	<i>EstacionesMareograficas.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
342	<i>Covid19Determinaciones.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE
343	<i>listado-establecimientos-salud-asentados-registro-federal-refes-201908.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
344	<i>cantidad-consultas-medicas-ambulatorias-subsector-oficial-argentina-anos-2013-2014.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
345	<i>farmacias-062018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
346	<i>establecimientos_farmacias_202009.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
347	<i>listado-establecimiento-farmacias-refar-201908.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
348	<i>gavade_20210104_120613.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
349	<i>establecimientos_droguerias_202009.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
350	<i>listado-establecimientos-droguerias-redro-201908.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
351	<i>droguerias-062018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE
352	<i>informacion-publica-dengue-zika-nacional-hasta-20191231.csv</i>	NO	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
353	<i>informacion-publica-dengue-zika-nacional-hasta-20181231.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE
354	<i>dengue-zika-nacional-201807.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
355	<i>recetas-medicamentos-esenciales.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
356	<i>recetas-medicamentos-esenciales-2003-2020.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
357	<i>rennya2020-09-31.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	CUMPLE
358	<i>renibase1.csv</i>	NO	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE
359	<i>tasa-mortalidad-infantil.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
360	<i>dicei-vacunas-srp-2009-2019.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
361	<i>recetas-medicamentos-escenciales-2003-2019.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE
362	<i>dicei-vacunas-srp-2009-2018.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	CUMPLE
363	<i>Covid19VacunasAgrupadas.csv</i>	NO	CUMPLE	CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE	NO CUMPLE
364	<i>cantidad-de-altas-mensual-por-ugl-2.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
365	<i>aportes_colectivoscd.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
366	<i>dat-ab-usuarios-2020.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
367	<i>operaciones-subedia-habil-promedio-grilla-hexagonal-modo-hora.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
368	<i>operaciones-de-viaje-por-periodo-modo-jurisdicion-grupo-tarifario-empresa-linea.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
369	<i>agp_detalle.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
370	<i>lineas-de-transporte.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
371	<i>volumenes-de-transporte-de-hidrocarburos-planilla-20.csv</i>	SI - EL ARCHIVO NO CUMPLE CON LA MISMA CANTIDAD DE COLUMNAS EN CADA UNO DE SUS REGISTROS								
372	<i>dat-ab-usos-2021.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
373	<i>matriz-tip-od-pax-2014.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
374	<i>matriz-tip-od-pax-2017.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
375	<i>matriz-tip-od-pax-2016.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
376	<i>empresas-de-transporte.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
377	<i>dat-ab-usuarios-2021.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
378	<i>trenes_pasajeros.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
379	<i>aeropuertos_detalle.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
380	<i>volumenes-de-transporte-de-hidrocarburos-planilla-21.csv</i>	NO	NO CUMPLE	CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	NO CUMPLE	CUMPLE	CUMPLE
381	<i>operaciones-de-viaje-por-periodo-modo.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



NRO DS	NOMBRE DEL DATASET	Bloqueante	Mét.1	Mét.2	Mét.3	Mét.4	Mét.5	Mét.6	Mét.7	Mét.8
382	<i>operaciones-de-viaje-por-periodo.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
383	<i>matriz-tip-od-pax-2015enrev.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								
384	<i>aportes_colectivosld.csv</i>	SI - EL ARCHIVO NO CUMPLE CON EL FORMATO CSV DE SEPARADOR (COMA)								



Anexo F: Detalle de los criterios de Validación

En esta sección se muestra la visión del enfoque utilizado para la deliberación que se realizó en base a si aplica o no cada una de las métricas, para los criterios de validación indicados.

ENFOQUE - CRITERIO DE PRECISIÓN

Aplica:

Este criterio toma en cuenta el dominio de las columnas, es decir, se analiza el conjunto de valores posibles que puede adoptar, por ejemplo: para una columna que es del tipo de dato numérico entero, su dominio válido son los valores que tengan sólo caracteres numéricos.

Para la reflexión sobre la métrica de Números Decimales (métrica 1) aplica ya que debe cumplir con un determinado rango sobre el formato del tipo número decimal (separador decimal, etc.). La métrica de caracteres inválidos (métrica 4), se hizo foco en que sus valores se deben encontrar en un dominio de caracteres determinados, los inválidos. Por otro lado, la Redundancia en el Dominio de una columna (métrica 5) aplica, ya que debe detectar que valores son los repetidos para una misma columna, es decir, su rango válido son determinados datos específicos (por ejemplo, valor de un campo es “organización” y se encuentra en otros valores de la misma columna, básicamente el dominio habilitado es “organización”, esto mismo se contempla para cada dato al realizar la detección de registro por registro del dataset). Para la métrica de Redundancia entre valores de Columnas de un mismo Registro (métrica 6), su deliberación fue similar a la anterior, pero en este caso, se contemplan valores comparando entre distintas columnas, no así en la misma. La métrica Estimación de ID (métrica 7) también aplica este criterio de análisis de dominio, ya que se deben rastrear los valores que contengan “id” o “id_” en los nombres de sus columnas, es decir, se buscan caracteres puntuales. Para la métrica de Campos Triviales (métrica 8) aplica, ya que se busca un único valor posible por columna, es decir, debe cumplir con ese dato puntual en el dominio. En cuanto a los aspectos generales, el tipo bloqueante para el archivo que posee doble carácter de comilla (aspecto 2), se busca comillas en el dominio, y para el aspecto bloqueante en base al nombre de columnas repetidas (aspecto 6), se analizan el primer nombre de columna como dato a buscar en el dominio (que son el resto de los nombres de columnas) y así sucesivamente con cada una.

No Aplica:

Para el resto de las métricas y aspectos en su análisis de dominio, no es necesario identificar valores específicos, por lo que no aplica el criterio de precisión.

ENFOQUE - CRITERIO DE EXACTITUD

Aplica:

Este criterio toma en cuenta un valor exacto que deben cumplir los datos de una columna. La métrica de registros duplicados (métrica 2), aplica, ya que se detectan que registros son exactamente iguales entre sí. La métrica de caracteres inválidos (métrica 4), aplica, ya que debe encuadrar con un conjunto de valores específicos, es decir, los delimitados para que sean considerados inválidos. Esto mismo sucede con las métricas de redundancia para el dominio de



una columna (métrica 5), la redundancia de valores de columnas de un mismo registro (métrica 6) en los que se busca valores determinados y el aspecto bloqueante de nombre de columnas repetidos (aspecto 6). Para la métrica de posibles campos triviales (métrica 8), ocurre algo similar, es decir, deben ser valores exactamente iguales entre sí para que sean contemplados como triviales. Para el aspecto bloqueante en el que el archivo posee doble carácter de comillas (aspecto 2), se busca un valor exacto, la doble comilla. Finalmente, el aspecto bloqueante sin formato CSV (aspecto 4), la extensión del archivo debe ser sí o sí CSV.

No Aplica:

Para el resto de las métricas y aspectos no se considera un análisis de valores exactos en sus valores, por lo que no aplica para el criterio de exactitud.

ENFOQUE - CRITERIO DE COMPLETITUD

Aplica:

Este criterio toma en cuenta que todos los valores deben estar cargados en el dataset. Una métrica que aplica es la de detección de datos incompletos (métrica 3), porque muestra directamente la cantidad de datos faltantes en caso de que existan. En lo que respecta a los aspectos, el único que aplica para este criterio es el tipo bloqueante que analiza la primera fila sin nombre de columnas (aspecto 5), ya que revela si falta alguna columna de título.

No Aplica:

Para el resto de las métricas y aspectos, no aplica este criterio, ya que pueden tener registros con valores incompletos, y sus validaciones internas no los detectarían, porque éste trabajo lo realiza la métrica 3, y en concepto de aspectos, el número 5.

ENFOQUE - CRITERIO DE INTEGRIDAD

Aplica:

Este criterio de calidad se enfoca en la estructura del dataset, coherencia de los valores y al concepto de la duplicación. Desde el concepto de duplicación, las métricas que aplican son las de registros duplicados (métrica 2), la redundancia en el dominio de una columna (métrica 5), la redundancia entre valores de columnas de un mismo registro (métrica 6) y la métrica de posibles campos triviales (métrica 8), ya que todas ellas son la base de la redundancia. Para el análisis de coherencia de valores, se aplica la métrica de caracteres inválidos (métrica 4), debido a que el contenido de las columnas, deben ser legibles para deducir un valor relativamente lógico, otra métrica orientada en este sentido es la estimación de ID (métrica 7), ya que detecta los casos que tienen un posible código identificador para un elemento. Desde el enfoque estructural del dataset, se aplican los 6 aspectos generales, esto se debe a su relación en forma directa con el análisis organizativo del archivo tanto interno como externo, por ejemplo: de manera externa con el tipo de extensión que sea CSV (aspecto 4), de forma interna, el resto de los aspectos.



No Aplica:

Para las métricas de números decimales (métrica 1) y la detección de datos incompletos (métrica 3) no aplica el criterio de integridad, ya que se orientan a enfoques de rangos válidos de valores decimales, y a datos faltantes.

ENFOQUE - CRITERIO DE CONSISTENCIA

Aplica:

Para este criterio de consistencia, se tienen en cuenta los conceptos de redundancia de datos para las métricas implicadas con la repetición: métricas 2, 5, 6, 8 y el aspecto número 6. En lo relativo a la coherencia de formatos para los datos, se aplican: métrica 1 (formato decimal) y métrica 7 (los valores deben contener la palabra "ID" a modo de formato condicional), y referente a los aspectos generales, el número 1 (aspecto de estructura en base a espacios en blanco), el número 2 (busca un conjunto de caracteres puntuales) y finalmente el aspecto 4 (por el tipo de formato CSV).

Para la coherencia lógica (libres de contradicción) del contenido del dataset, se aplica la métrica de caracteres inválidos (métrica 4) y para los aspectos generales, el número 2 (encuentra los delimitantes de contenido para su correcta identificación), 3 (necesario para discernir contenidos de las columnas) y 5 (conocer los nombres de las columnas), ya que sus valores deben ser legibles.

No Aplica:

La única métrica que no aplica es la 3, ya que se contemplan datos faltantes, no así aspectos de redundancia, coherencia o formatos.

ENFOQUE - CRITERIO DE RELACIÓN DE VALORES

Aplica:

Este criterio se orienta a la relación que puede existir entre los datos de las distintas columnas, básicamente aplica cuando hay relación o conexión alguna entre estos. Para la métrica de registros duplicados (métrica 2), aplica ya que se debe realizarse una comparación entre los registros, para la métrica de redundancia en el dominio de una columna (métrica 5) y la redundancia entre valores de columnas de un mismo registro (métrica 6), también aplican debido a que se deben buscar valores repetidos entre sí. Para la métrica de posibles campos triviales (métrica 8), aplica porque se localizan los valores redundantes para una misma columna. En lo que respecta a los aspectos generales, son aplicables los bloqueantes en base a la cantidad de columnas en cada registro (aspecto 3), ya que se debe asociar y recorrer cada registro para ver si cumple con la cantidad de columnas en cada uno. Finalmente, aplica el aspecto con nombres de columnas repetidas (aspecto 6), porque también se debe analizar nombre por nombre para comprobar que existe una redundancia.



No Aplica:

Para el resto de las métricas y aspectos generales no se aplica este criterio, ya que, para efectuar la validación de éstas, no utilizan un análisis que relacione los valores de distintas columnas o registros.

ENFOQUE - CRITERIO DE ESTRUCTURA/REPRESENTACIÓN

Aplica:

Este criterio se enfoca en la estructura del archivo del dataset, como así también en la representación de los valores, por lo que aplica para la métrica de números decimales (métrica 1), ya que se analiza si cumple un formato que represente el tipo decimal (separación decimal, etc.), por otra parte, aplica con todos los aspectos generales, debido a que se relacionan en un 100% con los conceptos que conforman la estructura del archivo para lograr una adecuada validación de calidad, por ejemplo, que se cumplan la misma cantidad de columnas en todos los registros (aspecto 3), espacios en blanco (aspecto 1), doble comillas para distinguir de un valor de otro y a su correcta asociación de columna (aspecto 2), el tipo de formato del archivo (aspecto 4), identificación de los nombres de columnas (aspecto 5) y nombre de columnas repetidas, para no confundir un dominio de valores que tal vez son correspondientes a una columna lógica, pero en realidad son de otra (aspecto 6).

No Aplica:

Para el resto de las métricas y aspectos no aplica el criterio, debido a que no poseen un análisis referido a su estructura o un determinado tipo de representación.

ENFOQUE - CRITERIO DE REDUNDANCIA

Aplica:

Este criterio apunta al análisis de repeticiones que pueden darse en los datos del dataset, por lo que las métricas que aplican este criterio son las orientadas a copias de datos, como ser: registros duplicados (métrica 2), métricas relativas a la redundancia (métrica 5, 6 y 8). En cuanto a los aspectos generales, aplica para el aspecto del tratamiento de nombres repetidos (aspecto 6).

No Aplica:

Para este criterio no se aplican las métricas y aspectos generales sin casos de duplicación o análisis de repetición de valores.



Acrónimos

ACM	Association for Computing Machinery.
ADD	Attribute Driven Design. Diseño Dirigido por Atributos.
AGA	Alianza para el Gobierno Abierto.
API	Interfaz de programación de aplicaciones.
APN	Administración Pública Nacional.
ASP .NET	Entorno para aplicaciones web desarrollado y comercializado por Microsoft.
BA	Buenos Aires
BID	Banco Interamericano de Desarrollo
CAETI	Centro de Altos Estudios en Tecnología Informática.
CAF	Corporación Andina de Fomento.
CDDQ	Dimensiones de la calidad de los datos (CDDQ) propuestas por Dan Myers en DQMatters.
CEPAL	Comisión Económica para América Latina y el Caribe
CKAN	Red Integral de Archivos de Conocimientos.
CSS	Cascading style sheets. Hojas de estilo en cascada.
CSV	Valores separados por coma (comma-separated values).
DataSF	Portal de datos abiertos de San Francisco.
DBF	Tipo de formato de archivo que permite almacenar datos en forma de tabla con los campos de datos fijos.
DKAN	Red de archivos de conocimiento de Drupal.
DOC	Formato del tipo Documento.
DOCX	Formato del tipo Documento a partir de Office 2007.
DSL	Lenguajes de Dominio Específico.
DSSI	Dirección de Seguridad y Sistemas de la Información.



DTA	El formato DTA se usa como archivos de datos de EZ-Forms para crear e imprimir todo tipo de documentos y formularios.
EITI	Iniciativa de transparencia de las industrias a nivel internacional (EITI, Extractive Industries Transparency Initiative).
ETL	Extracción, Transformación y Carga (Extract, Transform and Load).
FODA	Fortalezas, Oportunidades, Debilidades y Amenazas.
GEOJSON	Javascript Object Notation, es uno de los formatos GIS vectoriales basado en JSON que es muy rápido de analizar en máquinas virtuales Javascript.
GIGAPP	Grupo de Investigación en Gobierno, Administración y Políticas Públicas.
GIGO	Garbage In Garbage Out.
GODI	Global Open Data Index.
HEVDA	Herramienta de validación de datos abiertos.
HTML	Lenguaje de Marcas de Hipertexto, del inglés HyperText Markup Language.
IEEE	Instituto de Ingeniería Eléctrico y Electrónico.
IIS	Internet Information Services.
ILDA	Iniciativa Latinoamericana de Datos Abiertos.
INAP	Instituto Nacional de la Administración Pública
ISO/IEC	ISO (Organización Internacional de Normalización) e IEC (Comisión Electrotécnica Internacional) forman el sistema especializado para la normalización mundial.
JSON	JavaScript Object Notation, Notación de objeto de JavaScript.
KML	Keyhole Markup Language es un lenguaje de marcado basado en XML para representar datos geográficos en tres dimensiones.
MIT	Instituto de Tecnología de Massachusetts.
OD	Open Data. Datos Abiertos.
OD4D	Open Data for Development. Datos abiertos para el Desarrollo.
ODI	Open Data Institute.
ODIN	Open Data Inventory.
ODS	Archivo del tipo de format OpenDocument Spreadsheet (.ods).



OECD	Organisation for Economic Co-operation and Development. Organización para la Cooperación y el Desarrollo Económicos.
OG	Open Government. Gobierno Abierto.
OGD	Open Government Data. Datos gubernamentales abiertos.
OGP	Open Government Partnership. Alianza para el Gobierno Abierto.
OHDSI	Observational Health Data Sciences and Informatics.
ONTI	Oficina Nacional de Tecnologías de Información.
ONU	Organización de Naciones Unidas.
OWL	Web Ontology Language.
PDF	Portable Document Format, formato de documento portátil.
RAR	Es un formato de archivo privado, con un algoritmo de compresión sin pérdida utilizado para la compresión de datos y archivado.
RDF	Resource Description Framework.
RDFS	RDF Schema o Esquema RDF es una extensión semántica de RDF.
RISP	Reutilización de la Información del Sector Público.
SAV	Los archivos con la extensión SAV son creados como archivos de almacenamiento genérico por varios programas.
SHP	Shapefile (SHP) es un formato de archivo informático de datos espaciales.
SPARQL	SPARQL Protocol and RDF Query Language.
TIC	Tecnologías de la Información y de las Comunicaciones.
TXT	Un archivo de texto simple, texto sencillo o texto sin formato.
UIF	Unidad de Información Financiera.
URI	Identificador de recursos uniforme.
URL	Uniform Resource Location, localizador de recursos uniforme.
UTF-8	UCS Transformation Format 8 es la Codificación de caracteres Unicode e ISO 10646 que utiliza símbolos de longitud variable.
W3C	World Wide Web Consortium.
WPRDC	Centro de Datos Regional del Oeste de Pensilvania.



WWW	World Wide Web.
XLS/XLSX	XLS y XLSX son la extensión de archivo de la hoja de cálculo de Microsoft llamada Excel. La principal diferencia entre estas dos extensiones de archivo es que el XLS se crea en la versión de Excel anterior a 2007 mientras XLSX se crea en la versión de Excel 2007 y posteriores.
XML	Extensible Markup Language. Lenguaje de Marcado Extensible o Lenguaje de Marcas Extensible.
ZIP	Formato de compresión. Son archivos individuales que contienen uno o más archivos comprimidos.



Referencias Bibliográficas

- [AAD19] 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.
- [AAI21] Ministerio Público Fiscal, "*Agencia de Acceso a la Información Pública (AAIP)*", Disponible en: <https://www.mpf.gob.ar/agencia-de-acceso-a-la-informacion-publica-aaip/>, consultado en septiembre 2021.
- [ACE16] Acevedo, Sebastián; Dassen, Nicolás (2016), "*Innovando para una mejor gestión. La Contribución de los laboratorios de Innovación Pública*", Washington: Banco Interamericano de Desarrollo.
- [AEN21] AENOR, "*Marca AENOR N Sostenible*", Disponible en: <https://www.aenor.com/>, consultado marzo 2022.
- [AGE15] Argentina.gob.ar – Naciones Unidas, "*Asamblea General - Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*", Disponible en: https://www.argentina.gob.ar/sites/default/files/transformar_nuestro_mundo_documento_oficial_de_la_agenda_2030_original.pdf, consultado en septiembre 2021.
- [AGE21] Argentina.gob.ar – Argentina Unida, "*Agenda 2030 para el Desarrollo Sostenible*", Disponible en: <https://www.argentina.gob.ar/politicassociales/ods/institucional/agenda2030>, consultado en septiembre 2021.
- [AGN21] Auditoría General de la Nación, "*AGN*", Disponible en: <https://www.agn.gob.ar/>, consultado en septiembre 2021.
- [AGP21] Ministerio de Agricultura, Ganadería y Pesca (2021), "*Datos Agricultura, Ganadería y Pesca*", Disponible en: <https://datos.agroindustria.gob.ar/>, consultado en septiembre 2021.
- [ALB11] Alba Cuellar, D. (2011), "*Detección de registros duplicados entre dos archivos digitales*", Tesis de Maestría en Ciencias con orientación en Estadística Oficial, Disponible en: <https://ciemat.repositorioinstitucional.mx/jspui/bitstream/1008/239/2/TE%20370.pdf> consultado en septiembre 2021.
- [ALU10] Ramírez-Alujas, Á. V. (2010), "*Innovación en la gestión pública y open government (gobierno abierto): Una vieja nueva idea*", Innovation in Public Management and Open Government: An Old New Idea, Revista Buen Gobierno, (9).



- [ALU11] Ramírez-Alujas, Á. V. (2011), "Gobierno Abierto y Modernización de la Gestión Pública. Tendencias actuales y el (Inevitable) Camino que Viene-Reflexiones Seminales (Open Government and Modernization of Public Management: Current Trends and the (Inevitable) Way Forward-Seminal Reflections)", Revista Enfoques: Ciencia Política y Administración Pública, 9(15), 99-125.
- [ALU12] Alujas, Á. V. R., & Dassen, N. (2012), "Gobierno abierto: la ruta hacia una nueva agenda de reforma del Estado y modernización de la administración. Gobierno abierto y transparencia focalizada", 41. Edición: Nicolás Dassen y Juan Cruz Veyra, Banco Interamericano de Desarrollo (BID), pp. 41:71. Disponible en: <http://bit.ly/SWYNbR>, consultado en septiembre 2021.
- [AND21] Andino, "El portal redistribuible de datos de la República Argentina", Disponible en: <https://andino.datos.gob.ar/>, consultado en septiembre 2021.
- [AOK21] Open Knowledge Foundation (2020), "Global Open Data Index - Argentina", Disponible en: <https://index.okfn.org/place/ar/>, consultado en septiembre 2021.
- [AOU18] 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.
- [API21] datos.gob.ar, "API del Servicio de Normalización de Datos Geográficos de Argentina", Disponible en: <https://datosgobar.github.io/georef-ar-api/>, consultado en septiembre 2021.
- [ARC21] ArcGIS Hub, "ArcGIS Hub", Disponible en: <http://hub.arcgis.com/>, consultado en septiembre 2021.
- [ARI16] Arizo, I. (2016), "Métricas basadas en datos", Tesis de Maestría en Gestión de la Información, Universitat Politècnica de València.
- [ARR17] Arroyo Chacón, J. (2017), "La Innovación Abierta Como Pilar Del Gobierno Abierto", Open Innovation as a Pillar of Open Government, Revista Enfoques, 15(27), 13-41.
- [AUD21] Argentina unida, "Datos Argentina", Disponible en: <https://datos.gob.ar/>, consultado en septiembre 2021.
- [AVI14] Ávila Barrios, D. (2014), "El uso de las TICs en el entorno de la nueva gestión pública Mexicana", Andamios, 11(24), 263-288.
- [BAC21] Buenos Aires Ciudad, "Buenos Aires Data", Disponible en: <https://data.buenosaires.gob.ar/>, consultado en septiembre 2021.
- [BAR19] Chen, K., & Aitamurto, T. (2019), "Barriers for Crowd's Impact in Crowdsourced Policymaking: Civic Data Overload and Filter Hierarchy", International Public Management Journal, 22(1), 99-126.
- [BAR20] Barrera, M. A., Salgado, C., Peralta, M., Saldarini, J., & Claudio, C., "reutilización de datos públicos: propuesta de evaluación del grado de apertura de los datos en portales de infraestructura de datos espaciales", Jornadas de Ciencia y Tecnología 2020 "50 aniversario", 141.



- [BEL14] Beltrán Martínez, B. (2014), "*Minería de datos*", Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación.
- [BEM17] 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, Disponible en: <https://repository.ucatolica.edu.co/jspui/bitstream/10983/14642/1/Data%20Quality%20in%20OpenData.pdf>, consultado en septiembre 2021.
- [BIB21] Biblioguias at Biblioteca CEPAL, Naciones Unidas (2021), "*Datos Públicos Abiertos - De Gobierno Abierto a Estado Abierto*", Disponible en: <https://biblioguias.cepal.org/EstadoAbierto/datospublicos>, consultado en septiembre 2021.
- [BID20] BID – governate – Ideas innovadoras para mejores gobiernos (2020), "*Datos abiertos en gobiernos de América Latina y el Caribe: ¿cómo avanzamos?*", Disponible en: <https://blogs.iadb.org/administracion-publica/es/datos-abiertos-en-gobiernos-de-america-latina-y-el-caribe-como-avanzamos/>, consultado en septiembre 2021.
- [BID21] BID – Banco Interamericano de Desarrollo (2021), "*Gobierno Digital*", Disponible en: <https://www.iadb.org/es/dialogo-regional-de-politica/la-red-de-gobierno-electronico-de-america-latina-y-el-caribe>, consultado en septiembre 2021.
- [BOL12] Boletín Oficial del Gobierno de la Ciudad de Buenos Aires, "*Decreto 156/2012*", Disponible en: <https://boletinoficial.buenosaires.gob.ar/normativaba/norma/190097>, consultado en septiembre 2021.
- [BOL17] 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 en septiembre 2021.
- [BSA21] Buenos Aires DATA (2021), "*Historias con Datos*", Ciudad de Buenos Aires, Disponible en: <https://www.buenosaires.gob.ar/datosabiertos/historias>, consultado en septiembre 2021.
- [BYF16] Baylé, F. (2016), "*Detección de Villas y Asentamientos Informales en el partido de La Matanza mediante teledetección y sistemas de información geográfica*", Tesis de Maestría de la Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales, Disponible en: https://bibliotecadigital.exactas.uba.ar/collection/tesis/document/tesis_n6172_Bayle, consultado en septiembre 2021.
- [CAC17] Cáceres, M. H. (2017), "*Crowdsourcing y datos abiertos a través del uso de tecnologías móviles como instrumento de participación ciudadana en la ciudad de San José de Cúcuta*", Cuaderno Activa, 9, 37-49.
- [CAD19] 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.
- [CAF21] CAF, "*Banco de Desarrollo de América Latina*", Disponible en: <https://www.caf.com/>, consultado en septiembre 2021.



- [CAI15] Cai, L., & Zhu, Y. (2015), "The challenges of data quality and data quality assessment in the big data era", Data science journal, 14.
- [CAL12] Marotta, A., Vallespir, D., & Valverde, C. (2012), "Análisis de la calidad de datos en experimentos en ingeniería de software", In XVIII Congreso Argentino de Ciencias de la Computación.
- [CAL19] Calabrese, J., Esponda, S., Pasini, A. C., Boracchia, M., & Pesado, P. M. (2019), "Guía para evaluar calidad de datos basada en ISO/IEC 25012", In XXV Congreso Argentino de Ciencias de la Computación (CACIC), Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019.
- [CAT20] Nicolás, M., & Catachura, C. (2020), "Gobierno abierto: análisis de websites de datos abiertos gubernamentales en Argentina, Brasil y Paraguay", Instituições parceiras do INCT/PPED: UFRJ, UFF, UFRRJ, UFJF, UNICAMP e UERJ, 163.
- [CEI21] Comisión Europea (2021), "Inspirar Geoportal, Mejora del acceso a datos espaciales europeos", Disponible en: <https://inspire-geoportal.ec.europa.eu/linkagechecker.html>, consultado en septiembre 2021.
- [CEL21] Datos.gov.ar (2021), "Celdas vacías en filas para agrupar conceptos", Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/#celdas-vacias-en-filas-para-agrupar-conceptos, consultado en enero 2021
- [CEP21] CEPAL.org (2021), "Comisión Económica para América Latina y el Caribe", Disponible en: <https://www.cepal.org/es>, consultado en septiembre 2021.
- [CHA21] Government Challenge, Your Solutions (2021), "Challenge.gov", Disponible en: <https://www.challenge.gov/>, consultado en septiembre 2021.
- [CHU21] Gobierno del Chubut, "Datos Abiertos", Disponible en: <https://datos.chubut.gov.ar/>, consultado en septiembre 2021.
- [CIT20] Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020), "Citizen engagement with open government data. Transforming Government: People, Process and Policy".
- [CKA21] CKAN, "CKAN, the world's leading Open Source data portal platform", Disponible en: <https://ckan.org/>, consultado en septiembre 2021.
- [CMF21] Consejo de la Magistratura, "Fortalecimiento de las políticas de apertura y participación cívica en el Consejo de la Magistratura de la Nación", Disponible en: https://www.argentina.gob.ar/sites/default/files/fortalecimiento_de_las_politicas_de_apertura_y_participacion_civica_en_el_consejo_de_la_magistratura_de_la_nacion_1.pdf, consultado en septiembre 2021.
- [CMG21] Consejo de la Magistratura, "Poder Judicial de la Nación. Transparencia y Participación Ciudadana", Disponible en: <http://consejoabierto.pjn.gov.ar/>, consultado en septiembre 2021.
- [COD21] Gobierno de la Provincia de Córdoba, "Más Datos", Disponible en: <https://gestionabierta.cba.gov.ar/index.php/mas-datos/>, consultado en septiembre 2021.



- [COM19] Microsoft, “IDE de Visual Studio 2019”, Disponible en: <https://visualstudio.microsoft.com/es/vs/>, consultado en septiembre 2021.
- [CON16] Argentina.gob.ar, “Conformación del Consejo Federal para la transparencia”, Disponible en: <https://www.argentina.gob.ar/aaip/consejo-federal-transparencia/conformaci%C3%B3n>, consultado en septiembre 2021.
- [COR18] Corrêa, A. S., Corrêa, P. L. P., da Silva, F. S. C., & Carvalho, T. (2018), “An ADD-Oriented Software Architecture for Structuring Information to Open Government Data”, Journal of Computer Science, 14(5), 673-679.
- [COR21] Portal del Gobierno de Corrientes, “Corrientes”, Disponible en: <https://www.corrientes.gob.ar/>, consultado en septiembre 2021.
- [CRI19] Criado, J. I., & Pastor, V. (2019), “La Educación en los Planes de Acción de Gobierno Abierto”, RIESED-Revista Internacional de Estudios sobre Sistemas Educativos, 2(9), 279-300.
- [CRO20] Open Contracting Partnership, Colman R. (2020), “Women win one in four contracts in the Dominican Republic thanks to inclusive procurement reforms”, Disponible en: <https://www.open-contracting.org/2020/09/23/women-win-one-in-four-contracts-in-the-dominican-republic-thanks-to-inclusive-procurement-reforms/>, consultado en septiembre 2021.
- [CSV21] Datos.gob.ar (2021), “CSV”, Guía para la publicación de datos en formatos abiertos, Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/#csv, consultado en septiembre 2021
- [CUA21] Argentina.gob.ar, “Cuarto Plan de Acción de Gobierno Abierto 2019-2022”, Disponible en: <https://www.argentina.gob.ar/jefatura/innovacion-publica/plangobiernoabierto>, consultado en septiembre 2021.
- [DAR18] Darin, S. B. (2018), “La planificación estratégica para la cocreación de soluciones informáticas para el ciudadano en el marco de los principios del Gobierno Abierto”, Revista Publicando, 5(14 (1)), 107-115.
- [DCL20] DataCleaner (2020), “DataCleaner – Better data for better business decisions”, Disponible en: <https://datacleaner.org/>, consultado en septiembre 2021.
- [DCL21] data-cleaner, “data-cleaner 0.1.15 documentation”, Disponible en: <https://data-cleaner.readthedocs.io/en/latest/README.html#>, consultado en septiembre 2021.
- [DEC13] Boletín Oficial de la Ciudad de Buenos Aires, “Decreto 478/2013”, Disponible es: <https://boletinoficial.buenosaires.gob.ar/normativaba/norma/234859>, consultado en septiembre 2021.
- [DEC16] 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 en septiembre 2021.



- [DEC21] Naciones Unidas, “La Declaración Universal de Derechos Humanos” Disponible en: <https://www.un.org/es/universal-declaration-human-rights/#:~:text=Art%C3%ADculo%2019.,por%20cualquier%20medio%20de%20expresi%C3%B3n>, consultado en septiembre 2021.
- [DGE17] Datos.gob.es, reutiliza la información pública (2017), “Manual práctico para mejorar la calidad de los datos abiertos”, Disponible en: <https://datos.gob.es/es/documentacion/manual-practico-para-mejorar-la-calidad-de-los-datos-abiertos>, consultado en septiembre 2021.
- [DIM18] Conformed Dimensions of Data Quality (2018), “Annual Survey about Use of Dimensions of Data Quality”, Disponible en: http://dimensionsofdataquality.com/dims_survey, consultado en septiembre 2021.
- [DIP21] Datos Abiertos Diputados (2021), “Conjuntos de Datos”, Disponible en: <https://datos.hcdn.gob.ar/dataset>, consultado en septiembre 2021.
- [DIX20] Dixon, B. E., Wen, C., French, T., Williams, J. L., Duke, J. D., & Grannis, S. J. (2020), “Extending an open-source tool to measure data quality: case report on Observational Health Data Science and Informatics (OHDSI)”, *BMJ Health & Care Informatics*, 27(1), e100054.
- [DKA21] DKAN, “DKAN Open Data Platform”, Disponible en: <https://getdkan.org/>, consultado en septiembre 2021.
- [DSF21] ODI Open Data Institute, “The 2019 Data Skills Framework”, Disponible en: <https://theodi.org/article/open-data-skills-framework/>, consultado en septiembre 2021.
- [EIT21] Extractive Industries Transparency Initiative, “The global standard for the good governance of oil, gas and mineral resources”, Disponible en: <https://eiti.org/>, consultado en septiembre 2021.
- [EST21] Datos.gob.ar (2021), “¿Porqué es importante estandarizarlos”, Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-interoperables/#por-que-es-importante-estandarizarlos>, consultado en septiembre 2021
- [EUR19] European Data Portal (2019), “Open Data Maturity Report”, Disponible en: https://www.europeandataportal.eu/sites/default/files/open_data_maturity_report_2019.pdf, consultado en septiembre 2021.
- [EUR21] European Data Portal (2019), “Portal Europeo de Datos”, Disponible en: <https://www.europeandataportal.eu/es>, consultado en septiembre 2021.
- [FDT21] El futuro digital es de todos – Gobierno de Colombia MinTIC (2021), “Guía para el uso y aprovechamiento de Datos Abiertos en Colombia”, Disponible en: https://gobiernodigital.gov.co/623/articulos-9407_guia_datos.pdf, consultado en septiembre 2021.
- [FHP21] Fundación Huésped, “Prevención – Ciencia – Derechos – Fundación Huésped”, Disponible en: <https://www.huesped.org.ar/>, consultado en septiembre 2021.



- [FIT19] Fitriani, W. R., Hidayanto, A. N., Sandhyaduhita, P. I., Purwandari, B., & Kosandi, M. (2019), "*Determinants of Continuance Intention to Use Open Data Website: An Insight from Indonesia*", Pacific Asia Journal of the Association for Information Systems, 11(2).
- [FOR21] Portal Oficial Formosa || Gobierno, "*Datos Abiertos*", Disponible en: <https://www.formosa.gob.ar/datosabiertos>, consultado en septiembre 2021.
- [FRA16] Frank, M., & Walker, J. (2016), "*User Centred Methods for Measuring the Value of Open Data*", The Journal of Community Informatics, 12(2).
- [FRR17] Ferreira, R., Praia, V., Bonecini, F., Vieira, A., & Lopez, F. (2017, May), "*Platform of the brazilian csos: open government data and crowdsourcing for the promotion of citizenship*", In Anais do XIII Simpósio Brasileiro de Sistemas de Informação (pp. 17-24). SBC.
- [GAC21] Gobierno del Chubut, "*Gobierno Abierto del Chubut*", Disponible en: <http://www.gabierto.chubut.gov.ar/sitio/inicio>, consultado en septiembre 2021.
- [GAR11] Garriga-Portolà, M. (2011), "*¿Datos Abiertos? Sí, Pero De Forma Sostenible. El Profesional De La Información*", 20(3), Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2011/mayo/08.pdf>, consultado en septiembre 2021.
- [GBA21] Gobierno Abierto Buenos Aires Ciudad (2021), "*Gobierno Abierto*", Gobierno de la Ciudad Autónoma de Buenos Aires (CABA). Disponible en: <https://www.buenosaires.gob.ar/agendadetransparencia/gobierno-abierto>, consultado en septiembre 2021.
- [GCA21] Gobierno de la Provincia de Catamarca, "*Datos Catamarca*", Disponible en: <http://datos.catamarca.gob.ar/>, consultado en septiembre 2021.
- [GCH21] Gobierno de la Provincia de Chaco, "*Datos Abierto | Gobierno Abierto*", Disponible en: <https://gobiernoabierto.chaco.gob.ar/datos-abiertos-2/>, consultado en septiembre 2021.
- [GCO20] Gobierno de Colombia (2020), "*Guía de datos, LEILA – Librería de calidad de datos*", Unidad de Científicos de Datos Dirección de Desarrollo Digital, Disponible en: https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Digital/Big%20Data/2020/00_LEILA/LEILA_Presentacion.pdf, consultado en septiembre 2021.
- [GCO21] Gobierno de la Provincia de Córdoba, "*Datos Abiertos*", Disponible en: <https://datosgestionabierta.cba.gov.ar/>, consultado en septiembre 2021.
- [GEC21] Gobierno de España - Consejo Superior Geográfico (2021), "*Infraestructura de datos espaciales de España*", Europeo (Inspire), Disponible en: <http://staging-inspire-validator.eu-west-1.elasticbeanstalk.com/etf-webapp/home/index.html>, consultado en septiembre 2021.
- [GER21] Gobierno de Entre Ríos, "*Gobierno Abierto*", Disponible en: <https://www.entrierios.gov.ar/gobiernoabierto/>, consultado en septiembre 2021.



- [GES21] Guía de Estándares (2021), *“Calidad e Interoperabilidad de los datos abiertos del Gobierno de Colombia”*, Disponible en: <https://herramientas.datos.gov.co/>, consultado en septiembre 2021.
- [GGU21] Gualeguaychú | La ciudad sos vos, *“Bienvenida – Municipalidad de Gualeguaychú”*, Disponible en: <https://data.gualeguaychu.gov.ar/>, consultado en septiembre 2021.
- [GIG21] GIGAPP - Grupo de Investigación en Gobierno, Administración y Políticas Públicas (2021), *“Herramientas para mejorar la transparencia”*, Disponible en: <http://www.gigapp.org/index.php/component/djcatalog2/producer/1-herramientas-para-mejorar-la-transparencia>, consultado en septiembre 2021.
- [GIL17] Gill, M., Corbett, J., & Sieber, R. (2017), *“Exploring Open Data Perspectives from Government Providers in Western Canada”*, Journal of the Urban & Regional Information Systems Association, 28(1).
- [GIT21] GitHub (2021), *“Built for developers”*, Disponible en: <https://github.com/>, consultado en septiembre 2021.
- [GOB17] Gobernarte Ideas innovadoras para mejores gobiernos (2017), *“Análisis predictivo: Impulsar mejoras mediante el uso de datos”*, Disponible en: <https://blogs.iadb.org/administracion-publica/es/analisis-predictivo-impulsar-mejoras-mediante-uso-datos/>, consultado en septiembre 2021.
- [GOB21] Argentina.gob.ar, *“Gobierno Abierto”*, Disponible en: <https://www.argentina.gob.ar/jefatura/innovacion-publica/gobiernoabierto>, consultado en septiembre 2021.
- [GOE17] datos.gob.es (2017), *“Manual práctico para mejorar la calidad de los datos abiertos”*, Gobierno de España, Disponible en: https://datos.gob.es/sites/default/files/doc/file/manual_practico_para_mejorar_la_calidad_de_los_datos_abiertos_1.pdf, consultado en septiembre 2021.
- [GOM18] Gómez, C. E. J., & Roma, J. C. (2018), *“Análisis predictivo de datos abiertos sobre el uso turístico del servicio de alquiler compartido de bicicletas de Nueva York”*, Universidad Oberta de Catalunya, Master Universitario en Ciencia de Datos.
- [GOR15] Gorrochategui, Nora (2015), *“Innovación: gobernanza reflexiva y tecnologías de la información y las comunicaciones (TIC) en la gerencia pública Argentina”*, Trilogía Ciencia Tecnología Sociedad, vol. 7, no 13, p. 35-46.
- [GRA16] Grande, J. I. C. (2016), *“Las administraciones públicas en la era del gobierno abierto. Gobernanza inteligente para un cambio de paradigma en la gestión pública”*, Revista de estudios políticos, (173), 245-275.
- [GRA21] Graph Everywhere (2021), *“Principales indicadores para Calidad de Datos”*, Disponible en: <https://www.grapheverywhere.com/principales-indicadores-para-calidad-de-datos/>, consultado en septiembre 2021.
- [GSA21] Gobierno de Salta, *“Gobierno de Salta”*, Disponible <https://www.salta.gob.ar/>, consultado en septiembre 2021.



- [GSE21] Gobierno de Santiago del Estero, “Santiago, Gobierno de la Provincia de Santiago del Estero”, Disponible en: <http://www.sde.gov.ar/>, consultado en septiembre 2021.
- [GSF21] Gobierno de Santa Fe, “Datos Santa Fe”, Disponible en: <https://datos.santafe.gov.ar/>, consultado en septiembre 2021.
- [GTF21] Gobierno de Tierra del Fuego, “Gobierno de Tierra del Fuego, Antártida e Islas del Atlántico Sur, Datos Abiertos”, Disponible en: <https://gestiontransparente.tierradelfuego.gov.ar/tdf-data-2/>, consultado en septiembre 2021.
- [GTR21] Gobierno de Entre Ríos, “Transparencia”, Disponible en: https://www.entrerios.gov.ar/portal/index.php?codigo=36&item=pagina_textos&menu=menu&modulo=&accion=, consultado en septiembre 2021.
- [GTU21] Gobierno de Tucumán, “Conjunto de Datos”, Disponible en: <https://sep.tucuman.gov.ar/dataset>, consultado en septiembre 2021.
- [GUH21] datos.gov.ar, “Herramientas”, Disponible en: <https://datos.gov.ar/acerca/seccion/herramientas>, consultado en septiembre 2021.
- [GUI21] Datos.gov.ar (2021), “Guía para la identificación y uso de entidades interoperables”, Paquete de Apertura de Datos de la República Argentina, Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-interoperables/#guia-para-la-identificacion-y-uso-de-entidades-interoperables>, consultado en septiembre 2021.
- [GUP21] datos.gov.ar, “Guía para la publicación de datos en formatos abiertos”, Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/, consultado en septiembre 2021.
- [GUU21] Datos.gov.ar, “Guía para el uso y la publicación de metadatos”, Paquete de Apertura de Datos de la República Argentina, Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-metadatos/>, consultado en septiembre 2021.
- [HCD21] Honorable Cámara de Diputados de la Nación Argentina, “Diputados Argentina”, Disponible en: <https://www.diputados.gov.ar/>, consultado en septiembre 2021.
- [HLA19] Van Belle, J. P., & Hlabano, M. (2019, January), “Building Urban Resilience in the Face of Severe Drought through the Innovative Use of Open Data”, In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-7). IEEE.
- [IDA17] Datos Argentina – Paquete-apertura-datos (2017), “Guía para el uso y la publicación de metadatos”, Disponible en: https://paquete-apertura-datos.readthedocs.io/es/0.2.3/guia_metadatos.html, consultado en septiembre 2021.
- [IDE21] IdeaScale (2021), “Software de gestión de la innovación y la gestión de ideas”, Disponible en: <https://ideascale.com/es/>, consultado en septiembre 2021.
- [IGL19] Ibanez Gonzalez, L., Millard, I., Glaser, H., & Simperl, E. (2019), “An assessment of adoption and quality of linked data in European open government data”.



- [ILD20] ILDA – Open Data Barometer (2020), “Informe Regional Open Data Barometer Latin America and the Caribbean 2020”, Disponible en: <https://barometerlac.org/wp-content/themes/odbpress/reporte-ILDA-EN.pdf>, consultado en septiembre 2021.
- [INA17] INAP (2017), “XII Jornadas Nacionales del Sector Público: Nuestros aportes ante los desafíos del futuro”, Publicado el 31 de agosto 2017, Consejo Profesional de Ciencias Económicas de la Ciudad Autónoma de Buenos Aires. Disponible en: <https://archivo.consejo.org.ar/congresos/material/12publico/Straface.pdf>, consultado en septiembre 2021.
- [INA19] Curso INAP, Campus Virtual (2019), “Aspectos básicos de Innovación y Gobierno Abierto”, Modernización. Presidencia de la Nación.
- [IND20] Indart, C. (2020), “Datos abiertos de Investigación en Argentina. Análisis de su implementación en portales y repositorios. Obtenido de Master en sistemas de información digital”, Universidad de Salamanca: https://gredos.usal.es/bitstream/handle/10366/147099/TFM_SistemasInfoDigital_Indart_Camila_SI_85_2019-2020.pdf?sequence=1&isAllowed=y, consultado en septiembre 2021.
- [ISO08] 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>, consultado en septiembre 2021.
- [ISO12] ISO 25000 – Calidad de software y datos (2008), “ISO/IEC 25012:2008”, Disponible en: <https://iso25000.com/index.php/normas-iso-25000/iso-25012>, consultado en septiembre 2021.
- [ISO24] ISO 25024:2015 – Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of data quality (2015), “ISO/IEC 25024:2015”, Disponible en: <https://www.iso.org/standard/35749.html>, consultado en marzo 2022.
- [ISO40] ISO/IEC 25040 – Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) — Evaluation process (2011), “ISO/IEC 25040:2011”, Disponible en: <https://www.iso.org/standard/35765.html>, consultado en marzo 2022.
- [JET19] Jetzek, T., Avital, M., & Bjorn-Andersen, N. (2019), “The sustainable value of open government data”, Journal of the Association for Information Systems, 20(6), 6.
- [JUJ21] Gobierno Abierto Jujuy, “Datos Abiertos”, Disponible en: <http://www.gajujuv.gob.ar/portal-datos-abiertos/>, consultado en septiembre 2021.
- [KIT17] Buenos Aires provincia (2017), “Kit de Apertura Municipal”, Disponible en: <http://escueladefiscales.com/Kit%20de%20Apertura%20Municipal%202017%20-%20provincia%20de%20buenos%20aires.pdf>, consultado en septiembre 2021.
- [KUB18] Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018), “Comparison of metadata quality in open data portals using the Analytic Hierarchy Process”, Government Information Quarterly, 35(1), 13-29.



- [LAF18] Lafia, S., Turner, A., & Kuhn, W. (2018), *“Improving discovery of open civic data”*, In 10th International Conference on Geographic Information Science, (GIScience 2018). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- [LBN17] López Beltrán, N. E., & Mahecha Moyano, J. F. (2017), *“Prototipo de software para la evaluación de la calidad de datos abiertos”*, Tesis de Ingeniería de Sistemas, Universidad Católica de Colombia, Facultad de Ingeniería.
- [LCA20] 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 en septiembre 2021.
- [LEG16] La Legislatura de la Ciudad Autónoma de Buenos Aires sanciona con fuerza de Ley (2016), *“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 en septiembre 2021.
- [LEO19] 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).
- [LEY16] Infoleg - Información Legislativa, *“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>, consultado en septiembre 2021.
- [MAC18] Máchová, R., Hub, M., & Lnenicka, M. (2018), *“Usability evaluation of open data portals”*, Aslib Journal of Information Management.
- [MAG17] Magallón Rosa, R. (2017), *“Datos abiertos y acceso a la información pública en la reconstrucción de la historia digital”*, Disponible en: <https://e-archivo.uc3m.es/handle/10016/26128>, consultado en septiembre 2021.
- [MAN17] Manfredi-Sánchez, J. L. (2017), *“Horizontes de la información pública. El profesional de la información (EPI)”*, 26(3), 353-360, Disponible en: <http://www.elprofesionaldelainformacion.com/contenidos/2017/may/01.pdf>, consultado en septiembre 2021.
- [MAP20] Martínez, R., & Parkinson, C. (2020), *“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.
- [MAR20] Martínez, R., Rodríguez, R., & Vera, P. (2020), *“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.
- [MAR21] Martínez, R., Rodríguez, R. A., & Vera, P. M. (2021), *“Metrics proposal to measure the quality of governmental datasets”*, IEEE Latin America Transactions, 100(XXX), Disponible en: <https://latamt.ieeeer9.org/index.php/transactions/article/view/5642>, consultado en septiembre 2021.



- [MCO21] Municipalidad de Corrientes, “Portal de Datos Abiertos”, Disponible en: <https://datos.ciudaddecorrientes.gov.ar/>, consultado en septiembre 2021.
- [MED21] Medium (2021), “Innovación pública con ciencia de datos”, Disponible en: <https://eipdnp.medium.com/innovaci%C3%B3n-p%C3%BAblica-con-ciencia-de-datos-673cbbdc39d6>, consultado en septiembre 2021.
- [MEN21] Gobierno de Mendoza, “Datos Abiertos de Mendoza”, Disponible en: <http://datosabiertos.mendoza.gov.ar/>, consultado en septiembre 2021.
- [MET17] Open Knowledge Foundation Network (2017), “Methodology - Global Open Data Index”, Disponible en: <https://index.okfn.org/methodology/>, consultado en septiembre 2021.
- [MIC21] Microsoft (2021), “Data Quality Services”, Disponible en: <https://docs.microsoft.com/en-us/sql/data-quality-services/data-quality-services?view=sql-server-ver15>, consultado en septiembre 2021.
- [MIN20] Argentina.gob.ar, “Ministerio de Ciencia, Tecnología e Innovación de la República Argentina”, Disponible en: <https://www.argentina.gob.ar/ciencia>, consultado en septiembre 2021.
- [MIN21] Portal de Datos Justicia Argentina - Datos (2021), “Datos - Recursos acerca del funcionamiento del sistema de justicia de la República Argentina”, Disponible en: <http://datos.jus.gob.ar/dataset>, consultado en septiembre 2021.
- [MIS21] Gobierno de Misiones, “Datos Abiertos”, Disponible en: <http://www.datos.misiones.gov.ar/dataset>, consultado en septiembre 2021.
- [MIT02] The MIT Total Data Quality Management Program (2002), “MIT TDQM Program Highlight”, Disponible en: <http://web.mit.edu/tdqm/www/index.shtml>, consultado en septiembre 2021.
- [MMH17] Méndez Matamoros, J. H. (2017), “Mejoramiento de calidad en conjuntos de datos abiertos basado en la aplicación de métricas de consistencia lógica”, Tesis de Maestría en Ciencias de la Información y las Comunicaciones.
- [MMO16] 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>, consultado en septiembre 2021.
- [MOD18] Modernización del Estado (2018), “Serie de Investigaciones: Gobierno Abierto, Gobierno Digital y País Digital”, Instituto Nacional de la Administración Pública (INAP), Disponible en: https://www.argentina.gob.ar/sites/default/files/inap_serie_investigaciones_empleo_publico_modernizacion-estado-2018.pdf, consultado en septiembre 2021.
- [MON17] Montero, Gregorio (2017), “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. LC/PUB. 2017/9-P. p. 53-81, 2017.



- [MRO21] 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: Revista Ciencia e Investigación. ISSN 2528-8083, 6(2).
- [MSS20] Melo, C. A. H., & Sanabria, J. S. G. (2020), "Proposal for the Evaluation of Open Data Portals", Facultad de Ingeniería, 29(54), 1-20.
- [MUE18] Munte-Kunigami, A., & Serale, F. (2018), "Los datos abiertos en América Latina y el Caribe", Banco Interamericano de Desarrollo (BID), Disponible en: <https://clustertic.org/wp-content/uploads/2018/08/PublicacionDatosAbiertosFinal.pdf>, consultado en septiembre 2021.
- [NAC15] 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 en septiembre 2021.
- [NAS17] Naser, Alejandra; Ramírez Alujas, Álvaro (2017), "Plan de gobierno abierto: una hoja de ruta para los gobiernos de la región".
- [NAV20] Navarro, F. M. (2020), "¿Qué transparencia requiere el gobierno abierto?", Revista de Gestión Pública, 2(2), 303-333.
- [NAY18] 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>
- [NEU21] Gobierno de Neuquén, "Neuquén Open Data", Disponible en: <http://datos.neuquen.gob.ar/dataset>, consultado en septiembre 2021.
- [NOG21] Novagob (2021), "Ecosistema de Innovación Pública para Iberoamérica", Disponible en: <https://novagob.org/>, consultado en septiembre 2021.
- [NUL21] Datos.gov.ar (2021), "Valores nulos, desconocidos o en blanco en campos numéricos", Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/#valores-nulos-desconocidos-o-en-blanco-en-campos-numericos, consultado en septiembre 2021
- [OBV15] Open data - Biblioteca Virtual CLACSO (2015), "Open Data - Miradas y perspectivas de los datos abiertos", Obra Colectiva y Colaborativa, Disponible en: http://biblioteca.clacso.edu.ar/Argentina/unlar/20171117050559/pdf_1513.pdf, consultado en septiembre 2021.
- [OCD21] Open Contracting Partnership, "Herramienta de revisión de datos beta", Disponible en: <https://standard.open-contracting.org/review/>, consultado en septiembre 2021.
- [OCP21] Open Contracting Partnership, "The Open Contracting Data Standard", Disponible en: <https://www.open-contracting.org/data-standard/>, consultado en septiembre 2021.
- [OD421] OD4D, "Open Data for Development in Latin America and the Caribbean", Disponible en: https://www.cepal.org/sites/default/files/events/files/eliza_calza_od4d_p2.pdf, consultado en septiembre 2021.



- [ODB21] Open Data Barometer – World Wide Web Foundation, “*The Open Data Barometer*”, Disponible es: https://opendatabarometer.org/?_year=2017&indicator=ODB, consultado en septiembre 2021.
- [ODC21] The International Open Data Charter, “*Open Data Charter*”, Disponible en: <https://opendatacharter.net/>, consultado en septiembre 2021.
- [ODD21] ODD, “*Open Data for Development*”, Disponible en: <https://www.od4d.net/about.html>, consultado en septiembre 2021.
- [ODI16] Open Data Inception (2016), “*Open Data Inception - 2600+ Open Data Portals Around the World*”, Disponible en: <https://opendatainception.io>, consultado en septiembre 2021.
- [ODI21] 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 en septiembre 2021.
- [ODS14] Open Data Support (2014), “*Open Data & Metadata Quality*”, Disponible en: https://www.europeandataportal.eu/sites/default/files/d2.1.2_training_module_2.2_open_data_quality_en_edp.pdf, consultado en septiembre 2021.
- [ODS21] Argentina.gob.ar, “*Objetivos de Desarrollo Sostenible (ODS)*”, Disponible en: <https://www.argentina.gob.ar/ambiente/contenidos/ods>, consultado en septiembre 2021.
- [ODW20] Open Data Watch. (2020), “*Open Data Inventory (ODIN)*”, Disponible en: <https://odin.opendatawatch.com/>, consultado en septiembre 2021.
- [OECD20] OECD Better policies for better lives, “*Open Government Data*”, Disponible en: <http://www.oecd.org/internet/digital-government/open-government-data.htm>, consultado en septiembre 2021.
- [OEF18] de Oliveira, E. F., & Silveira, M. S. (2018, May), “*Open government data in Brazil a systematic review of its uses and issues*”, In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age (pp. 1-9).
- [OGD21] Open Government Data, “*The 8 Principles of Open Government Data*”, Disponible en: <https://opengovdata.org/>, consultado en septiembre 2021.
- [OGP12] Open Government Partnership, “*Argentina: Miembro desde 2012 | Plan de Acción 4*”, Disponible en: <https://www.opengovpartnership.org/es/members/argentina/>, consultado en septiembre 2021.
- [OGP21] Open Government Partnership, “*Sobre Open Government Partnership*”, Disponible en: <https://www.opengovpartnership.org/es/about/>, consultado en septiembre 2021.
- [OHD21] OHDSI (2021), “*Observational Health Data Sciences and Informatics*”, Disponible en: <https://www.ohdsi.org/>, consultado en septiembre 2021.
- [OKF20] Open Knowledge Foundation (2020), “*A fair, free and open future*”, Disponible en: <https://okfn.org/>, consultado en septiembre 2021.



- [OLA18] 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).
- [OPD12] 5 Open Data (2012), "5 Open Data", Disponible en: <https://5stardata.info/en/>, consultado en septiembre 2021.
- [ORA21] Oracle (2021), "Oracle Enterprise Data Quality", Disponible en: <https://www.oracle.com/ar/middleware/technologies/enterprise-data-quality.html>, consultado en septiembre 2021.
- [OSZ14] Oszlak, Oscar (2014), "Políticas públicas y capacidades estatales", Revista Forjando, vol. 3, no 5.
- [OVI13] E. Oviedo, JN Mazón y JJ Zubcoff (2013), "Hacia un modelo de calidad de datos para portales de datos abiertos", XXXIX Latin American Computing Conference (CLEI), Nanguata, 2013, pp. 1-8.
- [PAC21] Naciones Unidas Derechos Humanos, "Pacto Internacional de Derechos Civiles y Políticos", Adoptado y abierto a la firma, ratificación y adhesión por la Asamblea General en su resolución 2200 A (XXI), de 16 de diciembre de 1966, Disponible en: https://www.ohchr.org/sp/professionalinterest/pages/ccpr.aspx#:~:text=Art%C3%ADculo%2019&text=de%20sus%20opiniones._.2.,otro%20procedimiento%20de%20su%20elecci%C3%B3n, consultado en septiembre 2021.
- [PAM21] Home – Sitio oficial del Gobierno de La Pampa, "La Pampa, Gobierno en Acción", Disponible en: <https://www.lapampa.gob.ar/>, consultado en septiembre 2021.
- [PAN21] 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 en septiembre 2021.
- [PAS18] Pasini, A. C. (2018), "Modelos de evaluación de gobiernos abiertos, aplicado a los municipios de la provincia de Buenos Aires", In XXIV Congreso Argentino de Ciencias de la Computación, XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018).
- [PBA21] Gobierno de la Provincia de Buenos Aires, "Datos Abiertos PBA", Disponible en: <https://catalogo.datos.gba.gob.ar/>, consultado en septiembre 2021.
- [PCH21] Gobierno de la Provincia de Chaco, "Gobierno Abierto", Disponible en: <https://gobiernoabierto.chaco.gob.ar/>, consultado en septiembre 2021.
- [PIN04] Pinto, M. (2004), "Calidad y evaluación de los contenidos electrónicos", Disponible: <http://www.mariapinto.es/e-coms/calidad-y-evaluacion-de-los-contenidos-electronicos/>, consultado en septiembre 2021.
- [PIÑ08] Piñeiro, F. J. G. (2008), "Aspectos básicos de la calidad y de la gestión por procesos", Lurralde: Investigación y espacio, (31), 277-289.



- [PKI21] 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 en septiembre 2021.
- [POR11] Portals, D. (2011), "Data Portals", Disponible en: <http://dataportals.org/search>, consultado en septiembre 2021.
- [PWD21] PowerData (2021), "Calidad de Datos. Cómo impulsar tu negocio con los datos", Disponible en: <https://www.powerdata.es/calidad-de-datos#:~:text=Completitud%3A%20es%20el%20grado%20en,datos%20aparecen%20s%C3%B3lo%20una%20vez>, consultado en septiembre 2021
- [PYO18] Gascó-Hernández, M., Martín, E. G., Reggi, L., Pyo, S., & Luna-Reyes, L. F. (2018), "Promoting the use of open government data: Cases of training and engagement", *Government Information Quarterly*, 35(2), 233-242.
- [QUI16] Quintanilla, G., & Gil-García, J. R. (2016), "Gobierno abierto y datos vinculados: conceptos, experiencias y lecciones con base en el caso mexicano", *Revista del clac Reforma y Democracia*, (65), 69-102.
- [RIO21] Gobierno de La Rioja, "Dato Abierto", Disponible en: <https://web.larioja.org/dato-abierto>, consultado en septiembre 2021.
- [RIA19] 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), Disponible en: https://www.researchgate.net/profile/Ricardo_Tovar/publication/331966039_Valora_cion_de_factores_de_uso_de_los_datos_abiertos_de_gobierno/links/5c958a14299bf11169409c0c/Valoracion-de-factores-de-uso-de-los-datos-abiertos-de-gobierno.pdf, consultado en septiembre 2021.
- [RNE21] Gobierno de Río Negro, "Gobierno de Río Negro", Disponible en: <https://rionegro.gov.ar/>, consultado en septiembre 2021.
- [ROD18] Rodríguez, Exequiel (2018), "Laboratorios de innovación pública. Estado Abierto", *Revista sobre el Estado, la administración y las políticas públicas*, vol. 3, no 1.
- [ROY19] Royo-Montañés, S., & Benítez-Gómez, A. (2019), "Portales de datos abiertos. Metodología de análisis y aplicación a municipios españoles", *El profesional de la información*, 28(6).
- [RRL17] 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>, consultado en septiembre 2021.
- [RUI20] Ruijter, E., Grimmeliikhuijsen, S., van den Berg, J., & Meijer, A. (2020), "Open data work: understanding open data usage from a practice lens", *International Review of Administrative Sciences*, 86(1), 3-19.
- [SAJ21] Gobierno de San Juan, "Datos Abiertos San Juan", Disponible en: <https://www.datosabiertos.sanjuan.gob.ar/>, consultado en septiembre 2021.



- [SAL15] Saldivar, J. (2015), "*IdeaScale, una plataforma para fomentar la innovación abierta del sector público*", Ecosistema Urbano, Publicado el 17 de junio, Disponible en: <https://ecosistemaurbano.org/tag/crowdsourcing/>, consultado en septiembre 2021.
- [SAL21] Datos Abiertos del Ministerio de Salud - Datos (2021), "*Datasets*", Disponible en: <http://datos.salud.gob.ar/dataset>, consultado en septiembre 2021.
- [SAN17] Sangiambut, S., & Sieber, R. (2017), "*The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions*", Journal of the Urban & Regional Information Systems Association, 28(1).
- [SAN21] Santander CityBrain (2021), "*La comunidad donde construir juntos, un Santander más sostenible*", Disponible en: <https://www.santandercitybrain.com/>, consultado en septiembre 2021.
- [SAP21] Gobierno de Santa Cruz, "*Santa Cruz, Gobierno de la Provincia*", Disponible en: <https://www.santacruz.gob.ar/>, consultado en septiembre 2021.
- [SAT21] Gobierno de San Luis, "*Transparencia*", Disponible en: <http://www.sanluis.gov.ar/transparencia/>, consultado en septiembre 2021.
- [SCH12] 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.
- [SCN21] Gobierno de Santa Cruz, "*Nuestros Números*", Disponible en: <https://www.santacruz.gob.ar/nuestros-numeros>, consultado en septiembre 2021.
- [SEC21] Secretaría de Modernización. Presidencia de la Nación, "*Paquete de Apertura de Datos de la República Argentina*", Disponible en: <https://datosgobar.github.io/paquete-apertura-datos/guia-subnacionales/#1-que-son-los-datos-abiertos>, consultado en septiembre 2021.
- [SIN19] Sinif, L., & Bounabat, B. (2019, March), "*A general framework of smart Open Linked Government Data: Application in E-health*", In Proceedings of the 2019 2nd International Conference on Geoinformatics and Data Analysis (pp. 99-103).
- [SOC21] Socrata, "*Building something using data?*", Disponible en: <https://dev.socrata.com/>, consultado en septiembre 2021.
- [SON18] Son, J., Son, Y., & Kim, H. (2018), "*Towards a Semantic Discovery for Heterogenous Open Data by Interlinking Metadata of Datasets*", In International Semantic Web Conference (P&D/Industry/BlueSky).
- [SOS21] Sosteniblepedia, "*Gobierno Abierto*", Disponible en: https://www.sosteniblepedia.org/index.php?title=Gobierno_abierto, consultado en septiembre 2021.
- [SQL15] Geeks.ms (2015), "*Tratamiento de datos duplicados en SQL Server (1)*", Disponible en: <https://geeks.ms/lmblanco/2015/07/06/tratamiento-de-datos-duplicados-en-sql-server-1/>, consultado en septiembre 2021



- [TAL21] Talend (2021), “Open Studio for Data Quality”, Disponible en: <https://es.talend.com/products/data-quality/data-quality-open-studio/>, consultado en septiembre 2021.
- [TIP21] datos.gob.ar (2021), “Estándares según el tipo de Datos”, Disponible en: https://datosgobar.github.io/paquete-apertura-datos/guia_abiertos/#estandares-segun-el-tipo-de-datos, consultado en septiembre 2021.
- [TOM19] Tomorrow City, “Datos abiertos: ¿cómo pueden mejorar la participación ciudadana y la economía?”, Disponible en: <https://tomorrow.city/a/datos-abiertos-como-pueden-mejorar-la-participacion-ciudadana-y-la-economia>, consultado en marzo 2022.
- [UNE21] AENOR, “UNE 178301 Open Data”, Soluciones en certificación para las Smart cities: Confiabilidad de los datos, Disponible en: <https://www.aenorrepublicadominicana.com/certificacion/tecnologias-de-la-informacion/open-data#:~:text=La%20UNE%20178301%20de%20Open,los%20datos%20a%20los%20ciudadanos>, consultado en marzo 2022.
- [VAL21] Comisión Europea (2021), “INSPIRAR Validator”, Disponible en: <https://joinup.ec.europa.eu/collection/elise-european-location-interoperability-solutions-e-government/solution/inspire-reference-validator/news/test-beta-inspire-reference-validator-user-interface/>, consultado en septiembre 2021.
- [VEG18] Venegas Álvarez, M. A. (2018), “Estudio sobre el acceso abierto a datos gubernamentales: el caso de transparencia en Chile, Costa Rica y Uruguay”, Tesis de Maestría, Universidad De Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ciencias de la Computación, Disponible en: <http://repositorio.uchile.cl/bitstream/handle/2250/159564/Estudio-sobre-el-acceso-abierto-a-datos-gubernamentales-El-caso-de-transparencia-en-Chile-Costa-Rica-y-Uruguay.pdf?sequence=1&isAllowed=y>, consultado en septiembre 2021.
- [VEL19] Cadena-Vela, S., Fuster-Guilló, A., & Mazón, J. N. (2019), “Publicando datos abiertos considerando criterios de calidad”.
- [VEN17] Venegas, L., & Chong, W. (2017), “Recuperación en un repositorio de datos abiertos de datasets relevantes a un requerimiento de información expresado en lenguaje natural”, Doctoral dissertation, Universidad Católica de la Santísima Concepción, Disponible en: <http://repositoriodigital.ucsc.cl/handle/25022009/1384>, consultado en septiembre 2021.
- [VPJ17] 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).
- [W3B21] W3C – Brasil, Disponible en: <https://www.w3c.br/>, consultado en septiembre 2021.
- [W3C09] W3C (2009), “Publishing Open Government Data”, W3C Working Draft 8 September 2009, Disponible en: <https://www.w3.org/TR/gov-data/>, consultado en septiembre 2021.
- [W3C15] W3C (2015), “Modelo para datos tabulares y metadatos en la Web”, Disponible en: <https://www.w3.org/TR/tabular-data-model/>, consultado en septiembre 2021.



- [WEF18] WebFoundation.org (2018), “El barómetro de los datos abiertos. Edición de los líderes de la promesa al progreso”, Open Data Barometer, Disponible en: https://webfoundation.org/docs/2018/09/WF_ODB_Report2_Spanish_Screen.pdf, consultado en septiembre 2021.
- [WIE19] Wieczorkowski, J. (2019, June), “Barriers to Using Open Government Data”, In Proceedings of the 2019 3rd International Conference on E-commerce, E-Business and E-Government (pp. 15-20).
- [XHA20] Xhardez, V. (2020), “Datos abiertos en la Argentina: desafíos para la apertura y reutilización de datos públicos de gobierno”, Disponible en: http://www.ciecti.org.ar/wp-content/uploads/2020/12/DT22-Datos-abiertos_FINAL.pdf, consultado en septiembre 2021.
- [XIA18] Xiao, F., Lyon, L., Zou, N., & Gradeck, R. M. (2018), “Emerging Roles for Optimising Re-Use of Open Government Data”, The 13th International Digital Curation Conference takes place on 19–22 February 2018 in Barcelona, Spain.
- [ZAP20] Zapata, E., Scrollini, F., & Fumega, S. (2020), “¿Cuán abiertos están los datos públicos?”, El barómetro de datos abiertos de América Latina y el Caribe 2020, Disponible en: https://scioteca.caf.com/bitstream/handle/123456789/1710/%c2%bfCuan_abiertos_estan_los_datos_publicos_El_barometro_de_datos_abiertos_de_America_Latina_y_el_Caribe_2020.pdf?sequence=1&isAllowed=y, consultado en septiembre 2021.
- [ZNZ19] 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, 5(2), 12-18.