

Hacia un paradigma de calidad de Datos Abiertos adecuado al propósito de Gobierno Abierto

Sebastián Leonangeli¹, Rosa R. Maenza²

¹ Universidad Tecnológica Nacional, Facultad Regional Rosario
sleonangeli@gmail.com

² Universidad Tecnológica Nacional, Facultad Regional Rosario,
Universidad del Centro Educativo Latinoamericano
rmaenza@gmail.com

Resumen. El concepto conocido con el nombre de “Gobierno Abierto” o “Estado Abierto” nace como propuesta al desarrollo de una mayor participación civil en el proceso de toma de decisiones de gobernanza, promoviendo la transparencia entre el sector público y el privado. Este nuevo formato de interacción surge de la necesidad de abordar problemas cada vez más complejos, los cuales no pueden ser solamente resueltos desde la esfera política, sino que es necesario el articulado entre diversos actores heterogéneos. Dentro de esta iniciativa el trabajo propuesto en la línea llamada Datos Abiertos juega un papel fundamental puesto que, por un lado, la disponibilidad de los mismos los convierte en herramientas que la sociedad civil tiene a su disposición para participar en el proceso de toma de decisiones y también sirve como línea de base para establecer niveles de transparencia y rendición de cuentas por parte del gobierno. En este contexto, y con la finalidad de asegurar el cumplimiento de los fines propuestos es que se torna crucial el análisis de los Datos Abiertos evaluando su calidad. Para esto, este trabajo toma la bibliografía actual, aplica los frameworks de calidad de datos a determinados datasets y propone un marco teórico para nuevas mediciones, basado en uso y propósito más que en modelado y disponibilidad.

1 Introducción

El interés por el trabajo con Datos Abiertos dentro del paradigma de Gobierno Abierto (GA) viene avanzando los últimos años a un ritmo muy veloz, y así lo demuestra el crecimiento de publicaciones dedicadas a este tema. Los abordajes dentro de esta problemática, han puesto especial foco en desarrollar los marcos de trabajo adecuados para medir la calidad de los datos abiertos basados en estándares variados. Dentro de la extensa bibliografía al respecto, la mayoría se centra en analizar técnicamente la calidad de la plataforma y los datasets utilizados, concentrando los estudios en el modelado, la compatibilidad y la disponibilidad de los mismos. En contraste muy pocos estudios hacen foco en el uso concreto de los datos proporcionados,

en una evaluación de impacto, en su real aplicabilidad y en la funcionalidad que se está efectuando por parte de la sociedad civil.

Por tal motivo, este trabajo busca proporcionar un alcance diferente de evaluación, proponiendo analizar el posible grado de utilización por parte de los usuarios finales, es decir proponer medir en un sentido más amplio la calidad de los datos abiertos para evitar “afectar negativamente la participación civil” [1]. Bajo este supuesto, se ampliará el espectro del concepto de usabilidad web para Datos Abiertos en Gobierno Abierto (OGD siglas en inglés de Government Open Data) considerándolo como una cualidad que excede la mera facilidad de uso.

En palabras de Ubaldi “hacer push de datos no es suficiente para crear valor. Modelos robustos y estrategias para acercar a los actores deben ser también puestas en práctica para permitir el diálogo entre el sector público y los usuarios de OGD” [2]. En este sentido siempre es beneficioso recordar el sentido de adecuación para un propósito, tan extendido en la bibliografía sobre calidad. Si el propósito de los datos abiertos es el de colaborar a la iniciativa de Gobierno Abierto, es en base a la adecuación de ese propósito que deben ser medidos y calificados como tales.

Los Datos de Gobierno Abierto deben, por definición, poder ser empleados libremente, y esa es una característica intrínseca y esperable, pero de ningún modo esa idea debe restringir el concepto de la calidad de los datos proporcionados on-line. En especial, se espera que sean utilizados por la sociedad civil para diseñar soluciones a problemáticas y necesidades concretas, a la vez que permitan la toma de decisiones ciudadanas con conocimiento, pudiendo contar con una visión más global del contexto y considerando datos que no le son factibles de ser obtenidos de otra forma.

El fin último de OGD es, entonces, mejorar la calidad de vida de los ciudadanos dentro de una comunidad y por lo tanto optimizar las formas de interrelación entre los diferentes actores de la ciudadanía toda. En este marco se incluye además a las diferentes áreas gubernamentales que por momentos parecen compartimentos estancos donde la información no circula. Espacios de este tipo proporcionados vía web permiten que diferentes sectores de gobierno tengan a su disposición elementos que les ayuden a gestionar sus tareas de forma más eficiente. Es esta adecuación al propósito que se buscará estandarizar para acercarnos a un paradigma de calidad de OGD.

En el presente documento se explicitarán distintos métodos de calidad existentes en la bibliografía actual de la temática. Los mismos serán aplicados a los dataset del sitio de la Municipalidad de Rosario, y se tratará de demostrar cómo estas mediciones no arrojan información sobre el uso de OGD dentro del paradigma de GA. Trataremos, también, de ofrecer un marco apropiado y proponer análisis que hagan posible estimar la calidad teniendo en cuenta los factores arriba mencionados.

2 Usabilidad y Calidad en Datos Abiertos

Varios estudios contribuyen al estado del arte de esta temática, presentando frameworks para la medición de calidad de Linked Open Data (LOD) y específicamente OGD.

Uno de los trabajos más destacados y emblemáticos de usabilidad en la web fue llevado a cabo por Jacob Nielsen [3] [4] y su decálogo sobre reglas indispensables a ser tenidas en cuenta a la hora de publicar en la web:

1. visibilidad del estado del sistema ante algún proceso que realiza el usuario, mostrando que está aconteciendo en todo momento;
2. relación entre el sistema y el mundo real de forma que tenga un lenguaje familiar, entendible y siga las convenciones de orden natural y lógico;
3. control y libertad del usuario que le posibilite deshacer o rehacer alguna acción no deseada, teniendo fácilmente una salida que le permita volver a su situación inicial;
4. consistencia y estándares deben ser empleados por diseñadores antes que innovar con otros elementos que permitan mismas funcionalidades;
5. prevención de errores, facilitando la corrección de futuras acciones no deseadas, antes que un mensaje de aviso de error es preferible un buen diseño que prevenga la ocurrencia de problemas;
6. reconocimiento antes que recuerdo, las instrucciones de que acciones y opciones puede efectuar el usuario deben estar visibles o ser fácilmente recuperables en todo momento;
7. flexibilidad y eficiencia de uso, proporcionando diferentes opciones de interfaz dependiendo del tipo de usuario que se trate (experto o novato);
8. estética y diseño minimalista, los diálogos deben contener información relevante para aumentar la visibilidad relativa;
9. ayudar a los usuarios reconocer, diagnosticar y recuperarse de los errores, las indicaciones deben ser dadas en lenguaje claro y simple, indicando de forma precisa el problema y sugerir una solución constructiva al mismo;
10. ofrecer ayuda y documentación de forma fácil de encontrar, focalizada, concreta y con una serie de pasos a efectuar.

Con respecto a OGD particularmente, la mayoría de estos trabajos citan como fuente el enfoque por 5 estrellas de Berners Lee [5], quien propone 5 requerimientos para un dataset de tipo abierto:

1. Disponible en la web, bajo cualquier licencia abierta.
2. Disponible como dato estructurado entendible por máquina (formato de planilla de cálculo preferible a formato PDF)

3. Disponible en formato no propietario (formato CSV preferible al de planilla de cálculo de Excel)
4. Uso de estándares abiertos de W3C (RDF y SPARQL) y URI para identificación
5. Enlazar los datos a datos de otro proveedor para brindar contexto.

En mayor o menor medida, toda la bibliografía actual se basa en estos presupuestos y busca extenderlos. Vetro [1] reconoce la utilidad de estos requerimientos, pero sólo al nivel de codificación de los datos, pudiendo entonces llegar a un nivel de 5 estrellas y sin embargo publicar datos de baja calidad, ya que quedarían fuera de alcance dimensiones como precisión, completitud, consistencia y puntualidad en la publicación.

Un desarrollo similar hace Behkamal [6] cuando propone las 5 características inherentes de calidad en Linked Open Data: Consistencia, Completitud, Univocidad, Precisión sintáctica y Precisión semántica. A continuación, se describen brevemente cada uno de estos puntos:

- Precisión sintáctica puede ser definida como la validez de la estructura de un dataset, mientras que la precisión semántica se refiere a la relación existente entre el valor de un dato y su correspondencia en el mundo real.
- Univocidad es el grado en el cual un atributo carece de redundancias dentro de un dataset. Se entiende en este sentido que ontológicamente la representación de clases y propiedades es única e irreplicable dentro de un dataset.
- Consistencia implica que dos o más datos no presentan conflictos. En un sentido bastante amplio, se entiende que los datos no presenten contradicciones y sean coherentes en el contexto del dataset.
- Completitud se refiere a que toda la información relevante para presentar un concepto especificado en el diccionario de datos esté presente en el dataset.

Por su parte, Walker [7] también refiere a calidad de las 5 estrellas y avanza hacia un esquema de atributos en niveles similares a los expuestos en trabajos ya mencionados (capacidad de descubrimiento, granularidad, inteligibilidad, confianza y linkabilidad con otro dato), cada cual con métricas que fueron presentadas a dos conjuntos de usuarios y fueron rankeadas para un grupo de datasets en Low, Medium y High, de manera subjetiva.

Mientras tanto, Ubaldi [2] nos recuerda que el enfoque de 5 estrellas fue adoptado por Gran Bretaña en su “Open Data White Paper” [8] para medir la usabilidad de su OGD, y luego varios países de la Unión Europea siguieron ese rumbo. Sin embargo, en la reunión de Sebastopol [9] (California, Estados Unidos), organizada por Publi-cResource.Org, donde se reunieron 30 representantes de GA con el sponsor de la

Sunlight Foundation, Google y Yahoo, definieron 8 principios para considerar que un dato sea abierto:

1. **Completo:** Todos los datos públicos están disponibles. Los datos públicos no están sujetos a validación privada ni a limitaciones de seguridad o de privilegios.
2. **Primario:** Los datos son recolectados de la fuente, con el nivel más alto posible de granularidad, sin agregación ni modificación.
3. **Puntual:** Los datos son puestos a disposición lo más pronto posible con el fin de preservar el valor de los mismos.
4. **Accesible:** Los datos están disponibles al mayor rango posible de usuarios para el mayor rango posible de propósitos.
5. **Procesable por máquina:** Los datos están razonablemente estructurados para permitir su procesamiento automático.
6. **No discriminatorio:** Los datos están disponibles para todos, sin requerimiento de registro o ingreso mediante un login.
7. **No propietario:** Los datos están disponibles en un formato sobre el cual ninguna entidad tenga control exclusivo.
8. **Sin licencia:** Los datos no están sujetos a derechos de autor, patentes, o ninguna otra regulación.

A la vez, Ubaldi plantea que estos niveles establecen las condiciones de disponibilidad y de aptitud para el reuso, pero que “el verdadero valor de OGD se logra cuando hay interés y capacidad en reutilizar esos datos. Esto requiere entendimiento de la capacidad, así como también estrategias para estimular la demanda activa de datos útiles y motivar el uso y reuso de los datos” [2].

Según Maenza [10] para trabajar con datos abiertos es necesario corroborar que los datos a publicar tengan una finalidad útil a un grupo de la sociedad, teniendo en cuenta que en ocasiones pueden surgir otros usos diferentes a los que han sido considerados inicialmente.

Lee and Walk [11], por su parte, opinan que “Aún no hemos encontrado agencias gubernamentales que estén satisfechas con sus métricas para evaluar Gobierno Abierto y su apropiación por parte del público. Las métricas son esenciales para procurar iniciativas de gobierno abierto. Sin embargo, muchas de las métricas existentes tienden a medir procesos o cantidad en vez de resultados o impacto”.

Dawes [12] también apunta que “Intentar un aumento en sentido y reusabilidad de la información del sector público también implica representar y entregar los datos para que puedan ser fácilmente accesibles, buscados, procesados y enlazados a otros datos sin restricciones”, lo cual establece necesidades híbridas entre tecnología y voluntad política.

Por lo mencionado, la mayoría de las métricas propuestas en trabajos con marco empírico son en general de bajo nivel, no específicas de OGD.

Es esta bibliografía la que recorreremos en busca de métricas (en los casos en que los trabajos sean empíricos y no meramente teóricos) y las utilizaremos en la siguiente sección sobre determinados datasets para evaluar si arrojan información adecuada en las dimensiones de LOD y de OGD.

3 Caso de estudio, el portal de datos abiertos de la Municipalidad de Rosario

El encuadre de datos abiertos y su evolución puede ser analizado desde dos enfoques, uno el normativo y otro informático, sustentado en la implementación de proyectos concretos.

Respecto al enfoque normativo se puede hacer una reseña de los principales hitos destacando que en Argentina a partir del año 2000 se comienzan a realizar diversas iniciativas dentro de la temática de gobierno abierto, en particular la ley Nro 25326 es creada para establecer la Protección de Datos Personales. Pero es recién en el año 2014 que se establece como marco legal la sanción de la Ley N° 27078 de Tecnologías de la Información y las Comunicaciones donde se especifica el 1er Plan de Acción de Gobierno Abierto, posibilitando una normativa para el acceso a todos los habitantes de la nación a los servicios de la información y las comunicaciones en condiciones sociales y demográficas equitativas, procurando la accesibilidad y asequibilidad.

Actualmente a nivel nacional se está efectivizando el seguimiento del 3er Plan Nacional de Gobierno Abierto, con los 44 compromisos definidos para ser cumplidos entre 2017 y 2019, los cuales están siendo evaluados conforme a las especificaciones dadas por el organismo internacional denominado Alianza para el Gobierno Abierto (OGP en inglés Open Government Partnership). Cabe destacar que también ya se están discutiendo los lineamientos de trabajo para la co-creación del 4to Plan Nacional de Gobierno Abierto, siendo que el 8 de mayo de este año se llevará a cabo el 4to Encuentro Nacional de Gobierno Abierto, Innovación Pública y Tecnología en la ciudad de La Plata.

En la provincia de Santa Fe el Decreto N° 0692/2009 [13] constituye el primer proyecto provincial para cumplir los estándares internacionales respecto al acceso de la información pública como derecho colectivo.

En Rosario el Concejo Municipal sancionó la Ordenanza N° 9279 [14] el 12 de septiembre de 2014 creando el Programa Open Data Rosario, fundamentado en el derecho de todas las personas al acceso de la información pública.

Respecto al desarrollo de proyectos informáticos específicos, el gobierno de la ciudad de Rosario ha sido precursor de la comunicación web, dando origen en el año

1997 a uno de los primeros sitios municipales del país. Con el transcurso de los años el espacio de vinculación con la ciudadanía fue variando, siendo que en el año 2015 se implementó un desarrollo adaptado a las demandas de la Sociedad del Conocimiento. Así, empleando una nueva herramienta de gestión de contenidos, se trabajó más hacia el desarrollo de políticas de GA, efectuando una mayor interacción en las redes sociales, optimizando la visualización en móviles y creando una serie de plataformas acordes a la filosofía de transparencia y rendición de cuentas.

En particular, según el sitio Índice de Datos abiertos de ciudades argentinas <http://ar-cities.survey.okfn.org/>, puede verse a Rosario ubicada en el ranking 11, con un score de 67%, destacándose la apertura de los lugares públicos, del transporte público, del presupuesto y de los trámites administrativos.

En el año 2016 nace el proyecto que da por origen la creación del sitio datos abiertos, <https://datos.rosario.gob.ar/>, desarrollado en DKAN, un software Open Source basado en Drupal [15] con el objetivo de “seguir trabajando en la transparencia de la gestión e impulsar el desarrollo de nuevos servicios que beneficien a los ciudadanos”. Según el propio portal se establece: “Creemos que un ciudadano bien informado puede tomar mejores decisiones. La gestión y manipulación de datos abiertos permite al ciudadano adoptar un rol más activo, participando en los asuntos públicos, ayudando al municipio en la elaboración de las políticas públicas y creando propuestas de valor. La apertura contribuye a la igualdad de oportunidades en la comunidad y fomenta prácticas de participación y colaboración.”

Podríamos entonces determinar que lo anterior sería para OGD en general y para el caso de Rosario en particular, el propósito al cual deben adecuarse los datos que allí se publican, y que es en base a esas expectativas que deberían medirse.

Al momento de realizar este análisis el portal de Datos Abiertos de Rosario cuenta con datasets distribuidos en 14 temáticas, a saber:

1. Territorio y Población - 62 Datasets
2. Movilidad - 35 Datasets
3. Salud - 35 Datasets
4. Administración financiera - 34 Datasets
5. Participación - 10 Datasets
6. Actividad económica - 8 Datasets
7. Ambiente - 6 Datasets
8. Cultura - 6 Datasets
9. Seguridad - 6 Datasets
10. Recreación y deporte - 5 Datasets
11. Obras e infraestructura - 4 Datasets

12. Tecnología - 4 Datasets
13. Educación - 3 Datasets
14. Administración pública - 2 Datasets

Para el propósito de este trabajo, se someterán a evaluación diez datasets del portal, tomados de forma aleatoria. La lista de los datasets analizados se presenta a continuación:

1. Contenedores de residuos domiciliarios (Territorio y Población)
2. Paradas de transporte urbano de pasajeros (Movilidad)
3. Partos en efectores municipales (Salud)
4. Presupuesto 2016 - Recursos ejecutados (Administración Financiera)
5. Áreas barriales (Participación)
6. Catálogo de la videoteca (Cultura)
7. Centros comerciales a cielo abierto (Recreación y Deporte)
8. Cámaras de videovigilancia (Tecnología)

La lista de métricas y sus definiciones se encuentran a continuación en Tabla 1:

Tabla 1. Métricas

Dimensión	Descripción	Métrica
Compleitud	Análisis cuantitativo de celdas vacías y con valores aplicables al dominio de la columna	M1- Porcentaje de celdas completas
	Análisis cuantitativo de celdas vacías por fila	M2 – Porcentaje de filas completas
Conformidad	Indica el nivel de cumplimiento con este enfoque	M3 – 5 estrellas
Intelegibilidad	Análisis cuantitativo del formato entendible tanto para usuarios finales o para uso computacional	M4 – Porcentaje de columnas en formato comprensible
Exactitud	Análisis cuantitativo de valores correctos de acuerdo al dominio y al tipo de dataset	M5 – Porcentaje de celdas con información cierta
	Análisis cuantitativo del error en agregación y la escala de representación de datos (sólo	M6 – Porcentaje de certeza en la agregación

		aplicable en columnas agregadas)	
Exactitud semántica	Análisis cuantitativo de celdas sin valores atípicos con respecto a los rangos aceptables	M7 – Porcentaje de celdas fuera de rango	
Exactitud sintáctica	Análisis cuantitativo de celdas sin errores sintácticos	M8 – Porcentaje de celdas con errores sintácticos	
Univocidad	Análisis cuantitativo considerando uno menos el porcentaje de celdas con diferente URI pero mismas propiedades	M9 – Porcentaje de instancias redundantes	

Los resultados de correr las métricas arriba expuestas a los datasets seleccionados se muestra en la Tabla 2:

Tabla 2. Métricas con dataset de Municipalidad

	Dataset							
	1	2	3	4	5	6	7	8
M1	1	1	1	1	1	0.83	1	1
M2	1	1	1	1	1	0.92	1	1
M3	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
M4	1	1	1	1	1	1	1	1
M5	1	1	1	1	1	1	1	1
M6	1	1	1	1	1	1	1	1
M7	1	1	1	1	1	1	1	1
M8	1	1	1	1	1	1	1	1
M9	1	1	1	1	1	1	1	1

La inter-correlación de las métricas expuestas fue cubierta por los trabajos citados en la bibliografía por lo cual no lo repetiremos aquí y daremos por supuesta la validez de los modelos presentados.

A continuación en la Tabla 3, se presenta un comparativo entre los resultados obtenidos por los trabajos que han propuesto las métricas utilizadas y los resultados obtenidos para esas métricas en los datasets seleccionados para este trabajo.

Tabla 3. Relación entre los indicadores estándares y los resultados encontrados

Métrica	Bibliografía	https://datos.rosario.gob.ar
M1	0,94	0,97
M2	1	0,98
M3	0,6	0,6
M4	1	1
M5	1	1
M6	1	1
M7	0,52	1
M8	1	1
M9	1	1

Observamos entonces que, según las métricas presentes en la bibliografía actual, el sitio de OGD de la municipalidad de Rosario presenta métricas de Calidad de igual y en algunos casos superiores a los casos de estudio presentados en la bibliografía. Los datasets son proporcionados en distintos formatos como csv, gml, json o incluso pdf.

Llegado este punto, debemos preguntarnos, entonces, si estas métricas de calidad describen de manera positiva y sustantiva la propiedad de estos datasets de servir a los propósitos de Gobierno Abierto en términos de impacto.

Como primera medida se relevó con la Municipalidad de Rosario si actualmente existe algún tipo de utilización de métricas sobre sus datasets. Lo obtenido de las entrevistas es que el control de los datos en tanto consistencia, precisión sintáctica y semántica es de parte del dueño de los datos del dataset, quien revisa el contenido antes de enviarlo al sector de informática y luego de que es publicado. Por otro lado, el sector informática valida el formato y la visualización, busca si existen campos en blanco o registros en blanco. Luego, una encargada de comunicación define títulos comprensibles e intuitivos para el diccionario de datos.

Se evidencia que en este proceso no existe un flujo específico que contemple la medición de parámetros de calidad para los OGD de la Municipalidad de Rosario, con lo cual establecer una evaluación del impacto se convierte en una tarea compleja, sin embargo trataremos de aproximarnos a algunos lineamientos en la siguiente sección.

4 Propuesta del marco de trabajo

Debido a que se está proponiendo un nuevo framework, las métricas serán creadas en base a la bibliografía teórica no empírica existente, y empleando las encuestas realizadas a usuarios finales de OGD. Para este último punto hemos tomado encuestas realizadas por la OECD y por Open Knowledge International.

En el sitio del Global Open Data Index (GODI) [16] se explicita que “GODI intencionalmente limita su investigación a la publicación de Datos Abiertos. No observa otros aspectos del marco de datos abiertos como contexto, uso o impacto”.

Por otro lado, Lee y Kawalk [11] no pasan por alto el hecho de que la implementación de GA y OGD es incremental, y que cada etapa del proceso es importante. En su trabajo proponen un modelo que requiere cuatro etapas de implementación antes de considerar que se logran las expectativas de OGD. El “paso cero” es evaluar de qué datos dispone el gobierno para establecer una línea base. Con el primer paso se busca incrementar la transparencia de esos datos. El siguiente y segundo paso se trata de mejorar la participación activa en el uso de esos datos. El tercer paso se refiere a mejorar esa colaboración logrando un mayor impacto, mientras que el cuarto paso consiste en transformar esa participación en una aproximación ubicua de la sociedad civil a los datos abiertos.

También Ubaldi en este sentido expone que mientras se vayan logrando mayores niveles de participación y de impacto, el valor es cada vez mayor sobre todo para sectores vulnerables de la población.

Finalmente Robinson y Yu [17] subrayan un concepto crucial, que es que GA y OGD pueden tranquilamente existir el uno sin el otro: una creciente apertura de datos (lo que para Lee y Kawalk serían los primeros estadios de su modelo) no es indicador suficiente para medir responsabilidad sobre el impacto. Según sus palabras “un gobierno puede ser abierto en el sentido de ser transparente, incluso si no lo hace a través de la tecnología, y a la vez un gobierno puede proveer datos y aun así permanecer opaco e irresponsable”.

Vemos así que la medición del impacto causado por los Datos Abiertos es una preocupación existente en los estudios actuales y, que lejos de considerarse una métrica ajena al aspecto técnico de OGD, se considera que estos logran sus más altos estadios y despliegan todo su potencial, cuando se logra demostrar que la interacción es tal que con la apertura de datos se logra interpelar a los sectores más vulnerables de la sociedad civil, involucrándose de manera ubicua en la agenda del Estado.

La pregunta es, entonces, ¿qué metadata puede servir para arrojar luz sobre el uso e impacto de OGD?, realizando un análisis de los datos pueden explicitarse los siguientes comentarios.

El municipio de Rosario no publica la totalidad de la información censal, especialmente la referente a indicadores económicos, lo que podría llevar a interpretar cuáles serían los indicadores apropiados para entender lo que Ubaldi llama “sectores

vulnerables”, aquellos a quienes la apertura de datos les permitiría fomentar políticas públicas que ataquen sus problemáticas. La información que sí se publica (sobre todo de salud, nacimientos y defunciones) [17] se realiza a nivel distrito, con lo cual podemos pensar una posible dimensión de geolocalización del acceso a los datos abiertos como un metadato que nos permita comenzar a explorar OGD en el marco de GA.

El Centro de Información Económica (CIE), por su parte, publica un informe anual de facturación dividido en sectores de actividad, pero no hay información referente a los costos operativos ni a los salarios, con lo cual desde el punto de vista de la distribución no contamos con información pertinente. Sin embargo, desde el lado del consumo podemos pensar en el dispositivo mediante el cual se acceden a los datasets como un metadato que nos permita identificar, si no sectores económicos, al menos un inicio mediante el cual comenzar a establecer una línea de base de capacidad de consumo considerando el precio de mercado de los dispositivos.

Finalmente proponemos otra dimensión que se refiera al dataset observado, para completar tres dimensiones de medición de metadata y extraer métricas que puedan resultar significativas. En base a lo expuesto, se propone una distribución en cubo con dimensiones sobre geolocalización, dispositivo de acceso y dataset observado.

Nuevamente, se volvió a consultar a la Municipalidad de Rosario acerca de si actualmente se está recolectando algún dato referido a estas tres dimensiones, para realizar una evaluación empírica y aprobar o descartar el modelo. Lamentablemente este no es el caso, con lo cual no podemos demostrar empíricamente el valor del modelo ni someter las métricas propuestas a una línea de base que nos permita evaluar su aplicabilidad dentro del paradigma. Sin embargo, nuestra propuesta es comenzar a recolectar en principio estas mediciones para luego avanzar en cuestiones más concretas como encuestas a usuarios finales del sitio con el objetivo de cruzar los resultados con las iniciativas de los laboratorios de innovación, por ejemplo.

Por otro lado, en cuanto al proceso de estandarización respecto al modelo de carga de datos, se nos informó que elaboraron una guía de procedimientos la cual es empleada a nivel interno por un grupo reducido de empleados municipales que realizan la subida de los datos al sistema. Existe entonces un circuito más bien centralizado, donde las diferentes reparticiones se encargan de gestionar los datos y mantenerlos actualizados. Sin embargo, no fue tampoco explicitado si existe un análisis del ciclo de vida de los datos y si se tiene en cuenta algún tipo de tratamiento respecto a la caducidad de los dataset.

Cabe destacar que en España la RED de Entidades Locales para la Transparencia y Participación Ciudadana recientemente ha propuesto una guía [19] de cuarenta conjuntos de datos que consideran de utilidad para ser publicados por las entidades locales con el objetivo fomentar la reutilización y generar valor público con los datos. En Argentina desde la Secretaría de Modernización, dependiente de la Presidencia de la Nación están disponibles en el sitio una serie de herramientas y guías de buenas prácticas para la publicación, identificación y uso de entidades interoperables, para la publicación de metadatos y para la implementación [20].

5 Conclusiones

El paradigma propuesto por los Datos Abiertos a nivel de políticas públicas se transforma en una propuesta sin sentido en la medida que los datos no puedan ser reutilizados generando el real valor de impacto en la sociedad. Es necesario desarrollar políticas y prácticas sólidas e integrarlas en todos los organismos y agencias del Estado.

Coincidimos con las recomendaciones propuestas por el Informe dado por el Barómetro de Datos Abiertos – Edición Líderes [21]:

- Pasar a la apertura de datos por defecto con políticas en acción sostenibles a través de todo gobierno. Para esto es necesario elaborar planes claros, introducir directrices y procedimientos formales, proporcionar recursos financieros y humanos, empezar la divulgación proactiva de los datos, dar prioridad a la publicación de datos con un propósito y avanzar hacia la apertura por diseño.
- Construir y consolidar las infraestructuras de datos. Apostar al desarrollo de habilidades para la apertura de datos en los diversos niveles, promoviendo de forma proactiva el ecosistema. Desarrollar prácticas efectivas de gestión de datos, inversión en formación y actualización de la infraestructura técnica. Mejorar la calidad de los datos, desarrollar sistemas de datos listos para la apertura y crear conexiones entre departamentos y agencias.
- Publicando con un propósito. Trabajar en colaboración con los grupos cívicos y en conexión con grupos consultivos se garantiza una comunicación sostenida y con sentido que entienda qué datos se necesitan y cómo los mismos pueden ser empleados para mejorar los servicios y el propio gobierno.

Al realizar esta investigación tomando como caso análisis el sitio de datos abiertos proporcionado por la Municipalidad de Rosario hemos encontrado varios puntos que pueden ser destacados:

- La apertura de datos no es el fin último de OGD, y medir su calidad en términos de accesibilidad, frecuencia de actualización, completitud, y otras métricas de LOD nos darán, justamente, una medida de qué tan abiertos son los datos, pero no de qué tan útiles son como OGD.
- Esta brecha en la medición viene siendo estudiada y planteada por la bibliografía teórica, pero, hasta lo que hemos podido investigar, no existen aún trabajos empíricos al respecto.
- Tomando como ejemplo la Municipalidad de Rosario, hemos visto que sus datasets arrojan valores muy altos de calidad en tanto LOD, pero no se realiza ningún tipo de medición que ayude a evaluar el impacto, el punto focal del propósito de OGD.

Entendemos, entonces, junto con Ubaldi, que “la participación universal es esencial para evaluar el valor de OGD. Todos deberían ser capaces de usar, reusar y redistribuir la información sin discriminación de ninguna clase, ya que esas condiciones permiten una real participación universal”. Pensado en estos términos, debemos también preocuparnos por la brecha digital que no permite este tipo de uso y reuso de los datos por aquellos sectores de la sociedad civil que justamente se verían más beneficiados del acceso a los mismos.

En el corto plazo, Vetro propone “tomar en especial consideración el rol de los llamados intermediarios, como por ejemplo organizaciones civiles o gubernamentales que sean capaces de interpretar los datos y presentarlos efectivamente al público en general”.

Según Pintos y Marín [22] el sector infomediario es el que basa sus modelos de negocio en la reutilización de la información. Su operativa consiste en obtener información del sector público o no y actuando como intermediario crear un producto o servicio con un aporte de valor que es ofrecido a terceros. Según estos autores se pueden mencionar tres grupos significativos que pueden actuar como posibles agentes reutilizadores: las propias administraciones públicas, pues otros departamentos pueden beneficiarse de la disponibilidad de datos; las empresas ya sea para mejorar sus procesos internos, la calidad de sus productos o aportarles valor actuando como infomediario; la sociedad representada por ciudadanos individuales o asociaciones civiles que en variadas oportunidades han podido resolver diferentes problemáticas.

En tal sentido, se puede observar en el sitio de la Municipalidad un insipiente espacio de infomediarios que comienzan a hacer uso de los datos creando aplicaciones para móviles con el objetivo de proporcionar utilidad a la comunidad. Entre las aplicaciones que han sido desarrolladas pueden mencionarse: Mi bici, Tu bici del Ente de la Movilidad de Rosario (ETR); ¿Cuándo llega? de Globant; ¿Cómo llego? de Gamazzo; Rosario Turismo del Ente Turístico Rosario; Agenda Cultural Rosario de Movizen; Aimant de Mawape; Estacionamiento Medido de Teknoware; Comunidad donante Rosario, Multas de Tránsito Rosario y Precios Justos Rosario de Rosario Mobile Apps.

En particular, empoderar a la sociedad civil con OGD requiere estrategias y programas concretos que permitan construir la próxima generación de usuarios. Nuevas habilidades son requeridas que van más allá de la alfabetización en el uso eficiente de las tecnologías de la información y la comunicación. Se requiere no sólo contar con los datos abiertos sino una sociedad civil capacitada en la búsqueda y entendimiento de los mismos, capaz de abordar negociaciones efectivas con el sector público. En tal sentido las universidades toman un rol crucial trabajando junto al estado y la administración pública como integrantes del ecosistema educativo [23].

Por todo lo expuesto, son varios los desafíos propuestos para futuros trabajos, en la inmediatez pueden proponerse los siguientes:

1. Evaluar empíricamente el modelo de dimensiones planteado.

2. Trazar conexiones entre los resultados obtenidos y el impacto obtenido en políticas concretas basados en OGD o en laboratorios de innovación.
3. Evaluar estrategias para capacitar a la sociedad civil en usos de OGD que les puedan ser útiles y logren ser incorporados a la vida democrática en su totalidad.

Referencias

1. Vetro et al. Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*. 33. (2016)
2. Ubaldi, B. Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. Tech. Rep. OECD Publishing (2013)
3. Nielsen, J; Pernice, K. Técnicas de Eyetracking para usabilidad Web. Anaya multimedia. (2010)
4. Nielsen, J. Usabilidad. Diseño de sitios web. Prentice Hall. (2001)
5. Berners-Lee, T. Linked data-design issues. Tech. Rep. W3C (2006)
6. Behkamal, B; Kahani, M.; Bagheri, E. y Jeremic, Z. A Metrics-Driven Approach for Quality Assessment of Linked Open Data. *Journal of theoretical and applied electronic commerce research*, 9(2), 64-79. (2014).
7. Walker, F. y Thompson. Strengthening the evidence-base for open government in developing countries. *Open Government partnership*. (2015)
8. Open Data White Paper. Presented to Parliament by the Minister of State for the Cabinet Office and Paymaster General by Command of her Majesty. (2012)
9. Open Government Data Principles – Sebastopol. (2007). Disponible en http://resource.org/8_principles.html
10. Maenza, R. Del gobierno electrónico a las ciudades inteligentes. Instancias en un camino de ida. *Revista Virtual Arenotech Francia* (2015) Disponible en http://www.arenotech.org/wp-content/uploads/2016/03/6_innovacion-abierta.pdf
11. Lee, G. y Kwak, Y. H. An open Government Implementation Model: Moving to Increased Public Engagement. IBM Center for the business of Government. (2011)
12. Dawes, S. Stewardship and Usefulness: Policy Principles for Information-Based Transparency. *Government Information Quarterly*, Volume 27, Issue 4. (2010)
13. Decreto 0692/2009 regula el mecanismo de acceso a la información pública. Disponible en:

- <https://www.santafe.gov.ar/normativa/getFile.php?id=86930&item=41870&cod=9efe78dc961dc7997a47624505ddb7eb>
14. Ordenanza Nro 9279/2014 del Concejo Municipal. Disponible en: <https://www.rosario.gob.ar/normativa/verArchivo?tipo=pdf&id=103362>
 15. Drupal. Disponible en <https://www.drupal.org/>
 16. Global Open Data Index. Disponible en: <https://index.okfn.org/methodology/>
 17. Robinson, D. H, Yu. The New Ambiguity of Open Government. UCLA Law Review Discourse. (2012)
 18. Anuario de población y estadísticas vitales. Dirección General de Estadísticas. Disponible: https://www.rosario.gob.ar/ArchivosWeb/anuario_poblacion_estadisticas_vitales2017.pdf
 19. Datos abiertos FEMP 2019. 40 conjuntos de datos a publicar por las Entidades Locales. Federación Española de municipios y provincias. Red de entidades locales por la transparencia y participación ciudadana. Disponible en <http://femp.femp.es/files/3580-1937-fichero/DATOS%20ABIERTOS%20FEMP%202019.pdf>
 20. Herramientas. Sitio web de Secretaría de Modernización. Presidencia de la Nación. Argentina. Disponible en <https://datos.gob.ar/acerca/seccion/Herramientas>
 21. Open Data Barometer – Leaders Edition Barómetro de Datos Abiertos Informe de la promesa al progreso (2018) OpenData Barometer y World Wide Web Foundation. Disponible en <https://opendatabarometer.org/leadersedition/report/?lang=es>
 22. Pintos, C. y Marín, J.L. Open Data. Reutilización de la información pública. Netbiblio e Instituto Nacional de Administración Pública. (2013)
 23. Maenza, R. El reto de las universidades trabajando junto al Estado y la administración pública. Anales del XXII Congreso Internacional del CLAD sobre la Reforma del Estado y la Administración pública. Centro Latinoamericano de Administración para el Desarrollo. (2017)

