

Publicando datos abiertos considerando criterios de calidad

Susana Cadena-Vela¹, Andrés Fuster-Guilló², Jose-Norberto Mazón²

scadena@uce.edu.ec, fuster@ua.es, jnmazon@ua.es

¹ Universidad Central del Ecuador, Quito, Ecuador.

² universidad de Alicante, Alicante, España.

Pages: 295–308

Resumen: Los datos abiertos son considerados un mecanismo de democratización en el acceso a la información generada por organizaciones del sector público y para el desarrollo de servicios digitales generados por el sector infomediario. Sin embargo, esta tendencia ha presentado algunas barreras que van desde la calidad insuficiente de los datos publicados hasta la falta de mantenimiento de los portales donde se publican. Esta investigación realiza un análisis del estado de la cuestión 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. El marco de referencia fue validado utilizando un caso de estudio mediante la metodología de investigación en acción.

Palabras-clave: Datos abiertos, calidad de datos, reutilización información pública.

Publishing open data considering quality criteria.

Abstract: Open data is considered a mechanism of democratization in the access to information generated by public sector organizations and for the development of digital services generated by the infomediary sector. However, this trend has presented some barriers ranging from the insufficient quality of the data published to the lack of maintenance of the portals where they are published. This research carries out an analysis of the state of the issue in the field of open data, as well as of international standards and best practices in data quality in order to propose a frame of reference that enables the publication of open data with an adequate level of quality. The framework of reference was validated using a case study using action research methodology.

Keywords: Open data, data quality, reuse of public information

1. Introducción

Durante los últimos años, la cantidad y variedad de datos abiertos publicados por las administraciones públicas en todo el mundo ha sido creciente. Esta tendencia responde

a la constante evolución de la tecnología marcada fuertemente por la expansión de Internet (Ramos Simón et al., 2012). De hecho el uso de datos abiertos es considerado una oportunidad para el desarrollo económico, ya que su reutilización ha permitido el desarrollo de soluciones tecnológicas en áreas como biología, salud, geografía, turismo, entre otras (Pirozzi & Scarano, 2016).

El concepto de datos abiertos representa una oportunidad para que la sociedad pueda reutilizar e incluso, idealmente innovar, impulsando a que los gobiernos continúen proponiendo proyectos de apertura de datos con el fin de garantizar el acceso a información pública por parte de la ciudadanía (Garriga-Portolà, 2011), (Benson, 2008).

Esta tendencia ha presentado barreras como la insuficiente calidad de los datos que afecta negativamente al proceso de reutilización, ocasionando que la ciudadanía invierta tiempo en comprobar y aumentar la calidad de forma descentralizada y descoordinada (Vetrò et al., 2016), lo que conlleva que se abandonen los proyectos de portales de datos abiertos, tan pronto como son implantados para la publicación de datos.

La calidad de datos en formatos abiertos tiene varias dimensiones : (1) exactitud, es decir, datos que representan correctamente la entidad o evento del mundo real, (2) puntualidad, datos que representan la situación real y la agilidad en la publicación (3) consistencia, no contienen contradicciones, (4) exhaustividad, incluyen todos los elementos de datos que representan a la entidad o evento, (5) disponibilidad, accesibles ahora y con el tiempo, (6) conformidad, siguen las normas aceptadas, (7) credibilidad, son de fuentes confiables, (8) accesibilidad, son legibles por máquina, (9) relevancia, cantidad adecuada de datos (Quality, 2016).

Las cuatro primeras dimensiones (1-4) de calidad citadas, dependen totalmente de los sistemas de información de origen. Las últimas dimensiones (5-9) dependen de la gestión implementada en los portales de datos. El esfuerzo que hacen las organizaciones y administraciones públicas para publicar datos, se ve afectado cuando no existe una gestión adecuada y ocasiona que se desechen proyectos de datos abiertos al poco tiempo de ponerse en marcha (Benitez, Comber, & Huerta, n.d.). Desafortunadamente, las investigaciones actuales se han centrado exclusivamente en analizar la calidad de los datos ya publicados (Coleman, Hughes, & Perry, 2009), es decir los elementos de calidad del (5-9). De la literatura revisada no se evidencia aportes que profundicen la vinculación de las dimensiones (1-4) entre los sistemas de origen y el proceso de apertura de datos.

La presente investigación trata de cubrir este vacío mediante el diseño de un marco de referencia que considera la calidad de datos en procesos de publicación en abierto, posibilitando que se publiquen los datos con un nivel de calidad adecuado. Concretamente, este marco de investigación posee las siguientes características:

1. Se define un archivo maestro de apertura como mecanismo de vinculación de los datos abiertos con los datos en origen.
2. Se realiza una propuesta de indicadores para la validación de la calidad del conjunto de datos como un mecanismo de mejora continua previa a la publicación de datos abiertos.
3. Se aplica el marco de referencia a un caso de estudio.

El artículo se estructura de la siguiente manera: se describe la metodología de investigación llevada a cabo, se presentan un análisis de los trabajos relacionados y se representa el marco de referencia propuesto. Los resultados de la aplicación de nuestro marco de referencia se presentan en la sección 5 y, finalmente, la discusión de resultados y las conclusiones se detallan en las secciones 6 y 7, respectivamente.

2. Metodología de Investigación

La metodología de investigación llevada a cabo se basa en el proceso “investigación en acción” (Baskerville, 2018) y en la estructuración de conocimiento. Se utilizó un caso de estudio, para diagnosticar la problemática de los datos publicados, planteándose una hipótesis de trabajo en concordancia con la aplicación del marco de referencia y realizando varias iteraciones (según la metodología de investigación en acción descrita en la Figura 1) para mejorarlo. Se concluye con la presentación de los resultados. Las actividades que se llevan a cabo son las siguientes:



Figura 1 – Etapas de la investigación en acción

Articulación del problema. Se estableció la problemática de la calidad de los datos abiertos publicados.

Hipótesis dinámica. Se planteó lo siguiente: ¿la creación de un archivo maestro de apertura mejora la calidad de los datos publicados en formatos abiertos?

Diagnóstico. Se evaluó la calidad de los datos publicados en el portal de datos abiertos.

Diseño de la acción. Se realizó una planificación del nuevo proceso de apertura considerando el marco de referencia diseñado, se evaluó en cada una de las etapas con entrevistas al grupo interviniente en el proceso.

Pruebas y resultados. Se analizaron los resultados de cada etapa para realizar las mejoras respectivas.

2.1. Diagnóstico de la calidad de datos publicados en el portal

Se realizó un diagnóstico a través de entrevistas con los actores del proceso de publicación de los conjuntos de datos y una evaluación de los conjuntos de datos publicados. De estas actividades se obtuvieron los siguientes resultados:

- La evaluación de la calidad de los datos, se realiza sobre el conjunto de datos de docentes descargado del portal, que tiene un total de 2.321 docentes. De los indicadores aplicados se obtiene los resultados descritos en la Tabla 1.

Campos	Válidos	Blancos/nulos	% válidos
<i>Apellidos</i>	2.072	249	89,27%
<i>Nombres</i>	2.069	252	89,14%
<i>Correo_institucional</i>	1.180	1.141	50,84%
<i>Correo_personal</i>	1.031	1.290	44,42%
<i>Categoría</i>	2.057	264	88,63%
<i>Dedicación</i>	2.058	263	88,67%
<i>Relación_trabajo</i>	2.017	304	86,90%
<i>Horas de docencia</i>	1.480	841	63,77%

Tabla 1 – Resultados del diagnóstico sobre el conjunto de datos publicados

- Los registros duplicados fueron 263 que equivale a 12%.
- Se identifican datos incompletos, docentes sin carrera o facultad.

De las entrevistas a los responsables de la publicación se identifican las siguientes problemáticas:

- No se tiene estandarización en la relación laboral.
- La persona encargada de la publicación fue temporal.
- Los responsables de publicación realizaron un proceso de validación manual en el archivo entregado, lo que ocasionó alguna mejora, pero no en los datos de origen.
- No existía una normativa institucional actualizada.

2.2. Diseño de la acción

Una vez realizado el diagnóstico se estableció la hipótesis de trabajo como ¿la creación de un archivo maestro de apertura mejora la calidad de los datos publicados en formatos abiertos? y se planificaron las siguientes actividades:

- Capacitación del marco de referencia diseñado.
- Aplicación del marco de referencia.
- Evaluación de cada etapa con cada uno de los participantes.
- Análisis de resultados de calidad de datos obtenido.
- Acciones de mejora del marco de referencia.

Cada una de las actividades se convirtió en iteraciones que fueron evaluadas. Se realizó un total de tres iteraciones hasta lograr llegar a un porcentaje de calidad de datos aceptado por el grupo interviniente en el proceso.

2.3. Aplicación del marco de referencia

Para la aplicación del marco de referencia propuesto se realizaron las siguientes actividades:

- Organización y definición de requisitos. Se identificó dentro de la institución al responsable del portal de datos abiertos con un perfil de gestión que le permita coordinar las diferentes actividades para la publicación, como: Reuniones con los posibles reutilizadores para evaluar la estructura del conjunto de datos, los potenciales reutilizadores estaban conformados por docentes, estudiantes y administrativos. Organizar reuniones de levantamiento de requisitos. Definir al responsable de la generación del dato en este caso el administrador de la base de datos institucional.
- Archivo Maestro. Se definió el origen de los datos para el conjunto de datos a publicar.
- Preparación del dato. Una vez generado el conjunto de datos se preparó los siguientes archivos: conjunto de datos en cvs y pdf, diccionario de datos y metadatos.

3. Trabajos relacionados

Publicar datos en abierto presenta algunas barreras, una es la falta de calidad de los datos publicados, específicamente con problemas en la completitud, especificación de metadatos (descripción correcta del registro de datos), datos obsoletos, exactitud e interoperabilidad entre otros (Janssen, Charalabidis, & Zuiderwijk, 2012). Además, otros problemas relacionados con la apertura de datos se describen a continuación:

- Poca o ninguna estandarización de columnas o atributos, ni de unidades de medida, entre otros (Attard, Orlandi, Scerri, & Auer, 2015).
- Conjuntos de datos incompletos, datos faltantes, falta de atributos relevantes o datos incorrectos (Sadiq & Indulska, 2017).
- Poca variedad de datos publicados (Luna-Reyes, Bertot, & Mellouli, 2014).
- Falta de especificación del propósito de los datos, así como de las fuentes de datos y del responsable de los datos (Reis, Viterbo, & Bernardini, 2018).

En (Kubler, Robert, Le Traon, Umbrich, & Neumaier, 2016), se realiza un estudio de 44 portales de diferentes países, concluyéndose que el movimiento de apertura de datos propone oportunidades amplias que merecen ser exploradas. Sin embargo, existe un problema de calidad en los datos abiertos, que tiene un impacto en las iniciativas de gobierno abierto. Además, en este estudio, se considera que la falta de información completa sobre los datos publicados afecta directamente a su reutilización, ya que disminuye su impacto y afecta la sostenibilidad de un proyecto de datos abiertos.

Uno de los elementos que afecta la calidad es la diversidad de las fuentes de datos en los sistemas de origen (Heinrich, Klier, Schiller, & Wagner, 2018), por cuanto la

implementación de portales abiertos involucra la integración de datos provenientes de diversas fuentes siendo estas formarles o informales, lo que dificulta su publicación en abierto (Oviedo, Mazon, & Zubcoff, 2015).

La integración de datos provenientes de diferentes fuentes, según (Carretero, Gualo, Caballero, & Piattini, 2017), se aborda con el estructura de dato maestro que permite incluir definiciones para la recopilación de datos de diferentes orígenes (Benitez-Paez, Comber, Trilles, & Huerta, 2018).

De estas investigaciones citadas en la Tabla 2, se puede observar un resumen de la problemática encontrada y las características incluidas en la propuesta:

Problemática	Aspectos incluidos en el marco de referencia
No se tiene en cuenta la calidad de datos abiertos desde las fuentes de datos, sino que se intenta resolver los problemas de calidad de datos una vez publicados, lo que resulta costoso y complejo.	Se incluye la utilización de un archivo maestro que identifica la necesidad de información por parte del reutilizador, con un proceso de combinación e identificación del origen de los datos dentro de la organización. Para ello se propone un archivo maestro de apertura como una adaptación del estándar (ISO, 2009).
No existe en los datos publicados, una afirmación de nivel de precisión y completitud de los datos y mecanismo de reclamo por el incumplimiento de los niveles de calidad.	Se incluyen indicadores para evaluación de precisión y completitud y una propuesta en el archivo de metadatos para incluir los datos del responsable de la información (ISO, 2013).
Los datos se publican en abierto sin considerar de manera formal a los potenciales reutilizadores (usuarios).	Se incluye un análisis de los requisitos de calidad de los datos, que a menudo incluye encuestas a usuarios y administradores de datos en este caso a los reutilizadores e infomediarios.

Tabla 2 – Problemas encontrados en la apertura de datos

En la siguiente sección, se presenta el marco de referencia aplicado y los resultados alcanzados.

4. Descripción del marco de referencia para la apertura de datos

El marco de referencia se organiza en cinco etapas, diseñadas en correspondencia con el análisis de la bibliografía científica, así como con los estándares y buenas prácticas existentes en el ámbito de la apertura de datos, dando énfasis a los criterios de calidad en las características de completitud y coherencia para evaluar los resultados.

Organización y definición de requisitos. - Para iniciar un proyecto de datos abiertos, dando énfasis a la calidad, el primer paso es considerar a la calidad de los datos como un proceso. Por tanto, la institución que pretende abrir sus datos debe establecer una estructura organizacional que permita sostener este proceso. La estructura propuesta incluye a los usuarios potenciales, estos participarán en la etapa de levantamiento de requisitos y la evaluación una vez publicados los datos, junto con el responsable del portal y el responsable del dato. En los siguientes apartados se explica las funciones de cada uno de ellos.

Responsable del portal.- Será el encargado de explotar el valor crítico que los datos pueden proporcionar (Lee, 2014) y la relación con los potenciales reutilizadores, para establecer acciones de mejora. El responsable del portal deberá tener conocimientos de datos abiertos y con capacidades de negociación con los potenciales reutilizadores. Entre las principales actividades están identificar el grupo de reutilizadores e infomediarios, organizar las acciones para el levantamiento de los requisitos, definir el responsable del dato, diseñar un plan de apertura de datos, definir los formatos de publicación y diseñar planes de mejora de la calidad del dato publicado.

Responsable de la generación del dato. -Es el responsable del proceso que tiene en cuenta la definición de requisitos de datos, desde el momento en que se genera el dato internamente en la organización hasta que se publica. Debe tener conocimientos técnicos de bases de datos. Las actividades sugeridas son: recibir el levantamiento de requisitos por parte del responsable del portal, definir el origen del dato y el estado del dato en la organización (automatizado, manual), definir políticas, establecer la codificación de los datos, así como consideraciones de anonimización y normativa de privacidad interna.

Definir el archivo maestro de apertura.-Dentro de una organización, los datos maestros permiten relacionar todos los datos que tiene valor crítico en un solo archivo (Smith & McKeen, 2018), a este concepto se propone añadir criterios de apertura de acuerdo a las siguientes actividades:

- Indagación: identificar las aplicaciones y sistemas de origen, para los conjuntos de datos requeridos por los potenciales reutilizadores y definidos por la organización. En esta etapa se considerará las bases de datos y fuentes de datos externas (Vetrò et al., 2016). Esta actividad permite crear un archivo único que deberá contener la identificación del origen de los datos dentro de la organización, evaluar la calidad de los datos en origen, identificar la información susceptible de publicación considerando normativas de protección de datos e identificar la existencia o no de metadatos y diccionarios de datos de origen.
- Anonimización: definir la técnica de anonimización para el conjunto de datos creado (Smith & McKeen, 2018). Para este proceso es necesario contrastar los datos susceptibles de publicación con las normativas de protección de datos vigente, para definir dentro de la institución la técnica de anonimización correspondiente.
- Licencia: se puede identificar varios tipos de licencia, como por ejemplo, con reconocimiento que permite la explotación de los datos incluyendo una finalidad comercial; con reconocimiento pero sin uso comercial de los datos, es decir no se puede publicar con finalidades comerciales, es decir, no se permite un uso comercial de los conjuntos de datos de las posibles actividades comerciales (Attard et al., 2015).
- Formato de publicación: establecer los formatos de publicación considerando requisitos como disponibilidad como datos estructurados legibles por máquina; disponibilidad en formato no propietario; estándares abiertos de W3C (RDF y SPARQL) o URI (Ferrer-Sapena, Peset, & Alexandre-Benavent, 2011).

Estructurar el archivo de metadatos y diccionario de datos. - La definición de metadatos significa, crear un archivo que contenga la información de los conjuntos

de datos publicados, incluyendo información de origen, semántica del conjunto de datos publicado, información del porcentaje de calidad publicado. De la misma manera el diccionario de los datos publicados debe tener una descripción de cada uno de los campos incluidos en el conjunto de datos (Rafique, Lew, Abbasi, & Li, 2012). Para ello, se presenta una propuesta basada en ISO 8000, a la que se adicionan los criterios de apertura, los cuales son: estándar de anonimización, licencia de publicación y codificaciones estandarizadas. Adicionalmente se proponen indicadores de evaluación para validar los datos abiertos relacionados con completitud y precisión.

Para establecer el porcentaje de calidad en completitud y precisión se estableció, a manera de ejemplo, los indicadores definidos en la **Tabla 3**.

Descripción	Fórmula de Cálculo
<i>Compleitud</i>	Evaluación de valores nulos en los campos requeridos.
<i>Exactitud</i>	Porcentaje de datos nulos, blancos, incompletos, duplicados, datos sin formato, datos sin estándar, datos inconsistentes (Sadiq & Indulska, 2017).

Tabla 3 – Indicadores de calidad de datos

Esta actividad permitirá tener una descripción de los datos que están preparándose para ser abiertos. Se propone la estructura de datos descrita en la Tabla 4.

Nombre	Descripción
<i>Identificador</i>	Código único
<i>*Licencia</i>	Derechos de autor
<i>Fecha</i>	Fecha de publicación
<i>*Información de contacto</i>	Información de contacto correspondiente al responsable del portal
<i>Idioma</i>	Especifica el idioma en el que se almacena
<i>*Publicado</i>	Identifica el método de anonimización utilizado
<i>Propietario</i>	Entidad que genera el dato
<i>Relación</i>	Descripción con entidades que se relaciona
<i>*Aplicación</i>	Especifica el nombre de la aplicación que genera
<i>*Usuario del dato</i>	Especifica el usuario/grupo que requirió el dato
<i>Codificación</i>	Descripción de la codificación del dato
<i>*Origen</i>	Tipo de Origen del dato digital o manual

Tabla 4 – Estructura del diccionario de datos con características de apertura

Los campos marcados con (*) representan los que corresponden para apertura de datos, además es importante notar que el Diagrama de Flujo de Datos será el que se utiliza en la organización para la generación del datos y asociado a sus procesos internos.

Preparación del dato y publicación. -Previo a la publicación de datos, es necesario validar si los requisitos de los potenciales reutilizadores o infomediarios han sido cumplidos. Una vez que el dato maestro de apertura está listo se aplicará, nuevamente los indicadores propuestos en la **Tabla 1**. Si no se cumple el nivel de calidad requerido, se publica en el metadato la información correspondiente o se regresa al proceso de estructuración del dato en el archivo maestro de apertura. La publicación es la etapa final del marco de referencia donde el conjunto de datos se estructura en formatos abiertos (Bizer, Heath, & Berners-Lee, 2009). El formato de apertura fue seleccionado en la etapa donde se define el archivo maestro de apertura.

5. Resultados

Luego de la aplicación del marco de referencia con cada una de las etapas en la primera iteración se obtuvieron los siguientes resultados:

- El 30% de los campos incluidos en el conjunto de datos se encontraban en formatos manuales.
- Las codificaciones de la relación contractual y dedicación no estaban estandarizadas.

Con estos elementos se trabajó en el plan de publicación y mejora de los datos conjuntamente y se agregó la identificación del responsable de la generación del dato. Se define que el dato se genera en el Sistema de Recursos Humanos y que se alimenta con los datos del Sistema de Carga Horaria. Para esto se agrega al marco de trabajo la identificación de los responsables del dato y la necesidad de la estandarización de los catálogos dentro de la institución. Se agrega al marco de trabajo la necesidad de definir el estado del dato previo a la apertura.

En la segunda iteración se analiza el porcentaje de completitud y exactitud resultado del diagnóstico, definiéndose como meta llegar con un nivel de calidad mayor al 95% en todos los campos, por cuanto existen datos en los sistemas de origen inconsistentes como las nominaciones de los cargos. Además, los datos históricos tenían algunos elementos no estandarizados. En el marco de trabajo se agrega las atribuciones del responsable del portal de datos abiertos, por ejemplo, la relación con el administrador de la base de datos y el responsable del dato.

Del resultado de la aplicación del marco de referencia, se generaron varias normativas institucionales relativas a los datos de docentes, como estandarización de cargos y nominaciones.

En la tercera iteración se establece la necesidad de agregar al marco de trabajo el concepto de archivo maestro de apertura que contiene el origen del dato, su estado y una descripción de cada uno de los campos que constan en el conjunto de datos.

Luego de la aplicación de las tres iteraciones se evalúa la calidad de los conjuntos de datos previo a la apertura y se llega a los resultados descritos en la Tabla 4.

Campos	Válidos	Blancos/nulos	% válidos
<i>Apellidos</i>	2.321	0	100,00%
<i>Nombres</i>	2.321	0	100,00%
<i>Correo institucional</i>	2.312	9	99,61%
<i>Correo personal</i>	2.105	216	90,69%
<i>Categoría</i>	2.208	113	95,13%
<i>Dedicación</i>	2.321	0	100,00%
<i>Relación trabajo</i>	2.321	0	100,00%
<i>Horas de docencia</i>	2.321	0	100,00%

Tabla 5 – Resultados de la aplicación del marco de referencia

Se analizaron los registros duplicados detectando 41, que equivale a un 2%, que corresponden a docentes que dictan clases en varias carreras o varias facultades. En este aspecto, también se agregó una política institucional tal que el docente pertenece a la facultad y carrera donde mayor carga horaria tiene.

6. Discusión

De los resultados obtenidos en la aplicación del marco de referencia para la apertura de datos, se valida la implicación directa entre la calidad de datos en origen con la calidad de datos publicados. Consecuentemente, para favorecer la sostenibilidad no deben ser procesos aislados, y se genera un archivo maestro de apertura basado en algunos resultados presentados en la investigación de (Heinrich et al., 2018).

La organización y definición de requisitos, permite mejorar la calidad de datos publicados por cuanto, al identificar a los potenciales reutilizadores facilita los procesos de levantamiento de requisitos de publicación, directamente con los consumidores de la información y la evaluación de la calidad, esto fue propuesto en la investigación de (Conradie & Choenni, 2014), mientras que en el presente trabajo se concluye que la participación de los potenciales reutilizadores es fundamental, caso contrario se publica lo que la institución defina.

El responsable del portal es el interlocutor de todo el proceso en el caso en estudio, por tanto se logró establecer en la segunda iteración las acciones que debe cumplir del responsable del dato, concluyendo en características similares con (CDO, Chief Data Officer) y (Lee, 2014), adicionando conocimientos de anonimización de datos y formatos de apertura.

Al realizar el archivo maestro de apertura, especialmente en la definición de los orígenes de datos, se mejora la calidad de los datos para la publicación en abierto, ya que se pudo

establecer un solo origen del dato. También se impactó a los procesos internos de la institución. Entre algunos ejemplos se puede citar: establecer claramente el catálogo de dedicación y categoría del docente, el cual no existía desde la Dirección de Recursos Humanos implantar una política institucional de uso del correo electrónico para el acceso a este servicio. Es importante también notar que identificar el responsable del dato permite mejorar los datos en origen y diseñar las acciones de mejora continua.

También se define al responsable de la creación del dato, con el propósito de institucionalizar la mejora de los procesos que deben ser automatizados. Esto facilitó visualizar nuevos requerimientos de sistemas de información y mejorar el diagrama de flujo de datos interno para poder generar el diccionario de datos para el portal en abierto.

Finalmente, se prepara el dato para publicación y evaluación. La mejora significativa de los datos se debió a que previo a la publicación se evaluó la calidad de los datos en las dimensiones definidas, hasta lograr un nivel de calidad superior al 90%. Esto incrementó el tiempo para la publicación, pero simultáneamente permitió la mejora de los datos en origen acompañados de la calidad en completitud y exactitud, demostrando con no pueden ser procesos aislados, ni con una temporalidad definida.

7. Conclusiones

En este trabajo se presenta un marco de referencia para la apertura de datos basado en estándares y buenas prácticas, que permite considerar criterios de calidad. En concreto, la aplicación del marco de referencia permite tener en cuenta la calidad de los datos en origen para conseguir una mejora en la calidad de los datos abiertos, como un proceso de mejorar continua.

Considerando que los datos publicados tienen su origen en aquellos que se gestionan en la organización, se debe incluir consideraciones de apertura de datos en formatos abiertos (y su calidad) desde el inicio de los procesos de automatización.

El marco de referencia presentado en este artículo, establece cada una de las actividades para definir un proceso de análisis de requisitos relacionado con los datos abiertos, además propone la creación del archivo de datos maestros de apertura (mediante la adaptación del estándar ISO 8000), con su correspondiente metadatos y diccionario de datos con las características necesarias para la publicación en abierto. Se presenta un conjunto de responsabilidades y roles en un proceso de apertura.

Tal y como se ha detallado en la validación, utilizar el marco de referencia propuesto ocasiona mayor inversión de tiempo en la preparación de los datos previo a la publicación, sin embargo, permite la mejora de las estructuras internas que facilitarán procesos de publicación posteriores con criterios de calidad.

Las limitaciones de la investigación se orientaron a un solo caso de estudio, sin embargo, la propuesta diseñada responde a los requerimientos levantados de la literatura científica publicada y estándares reconocidos como ISO, por tanto, podría ser aplicado a otras instituciones y se realizó mediante iteraciones que mejoraron el mismo. Los elementos citados generan un nuevo reto para el siguiente trabajo en incluir elementos de gobierno de datos.

Por tanto, para trabajos futuros se está implementando con este primer modelo una estructura de gobierno de datos para sostener los trabajos de apertura, con validación de expertos en cada una de las etapas propuestas.

Referencias

- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, 32(4), 399–418. <https://doi.org/10.1016/j.giq.2015.07.006>
- Baskerville, R. L. (2018). Investigating Information Systems with Action Research. *Communications of the Association for Information Systems*, 2(October). <https://doi.org/10.17705/1cais.00219>
- Benitez-Paez, F., Comber, A., Trilles, S., & Huerta, J. (2018). Creating a conceptual framework to improve the re-usability of open geographic data in cities. *Transactions in GIS*, 22(3), 806–822. <https://doi.org/10.1111/tgis.12449>
- Benitez, F., Comber, A., & Huerta, J. (n.d.). IMPROVE THE REUSABILITY OF OPEN How much data is generated every minute ?
- Benson, P. (2008). ISO 8000 the International Standard for Data Quality. MIT Information Quality Symposium, July 16-17(2008), 1–4.
- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Carretero, A. G., Gualo, F., Caballero, I., & Piattini, M. (2017). MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. *Computer Standards and Interfaces*, 54(September 2016), 139–151. <https://doi.org/10.1016/j.csi.2016.11.008>
- Coleman, D. W., Hughes, A. A., & Perry, W. D. (2009). The role of data governance to relieve information sharing impairments in the federal government. 2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009, 4, 267–271. <https://doi.org/10.1109/CSIE.2009.630>
- Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31(SUPPL.1), S10–S17. <https://doi.org/10.1016/j.giq.2014.01.003>
- Ferrer-Sapena, A., Peset, F., & Aleixandre-Benavent, R. (2011). Acceso a Los Datos Públicos y Su Reutilización: Open Data y Open Government. *El Profesional de La Informacion*, 20(3), 260–269. <https://doi.org/10.3145/epi.2011.may.03>
- Garriga-Portolà, M. (2011). ¿Datos Abiertos? Sí, Pero de Forma Sostenible. *El Profesional de La Informacion*, 20(3), 298–303. <https://doi.org/10.3145/epi.2011.may.08>
- Heinrich, B., Klier, M., Schiller, A., & Wagner, G. (2018). Assessing data quality – A probability-based metric for semantic consistency. *Decision Support Systems*, 110(October 2017), 95–106. <https://doi.org/10.1016/j.dss.2018.03.011>

- ISO. (2009). Chapter-6-Metadata-Management-for-MDM_2009_Master-Data-Management.
- ISO. (2013). International Standard Iso ISO/IEC 25024. 2015, 2015.
- Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4), 258–268. <https://doi.org/10.1080/10580530.2012.716740>
- Kubler, S., Robert, J., Le Traon, Y., Umbrich, J., & Neumaier, S. (2016). Open Data Portal Quality Comparison using AHP. *Proceedings of the 17th International Digital Government Research Conference on Digital Government Research*, 397–407. <https://doi.org/10.1145/2912160.2912167>
- Lee, Y. (2014). ESD Working Paper Series A Cubic Framework for the Chief Data Officer : Succeeding in a World of Big Data ESD-WP-2014-34. (November).
- Luna-Reyes, L. F., Bertot, J. C., & Mellouli, S. (2014). Open Government, Open Data and Digital Government. *Government Information Quarterly*, 31(1), 4–5. <https://doi.org/10.1016/j.giq.2013.09.001>
- Oviedo, E., Mazon, J. N., & Zubcoff, J. J. (2015). Quality and maturity model for open data portals. *2015 Latin American Computing Conference (CLEI)*, 1–7. <https://doi.org/10.1109/CLEI.2015.7360011>
- Pirozzi, D., & Scarano, V. (2016). Support Citizens in Visualising Open Data. *2016 20th International Conference Information Visualisation (IV)*, 271–276. <https://doi.org/10.1109/IV.2016.45>
- Quality, D. (2016). Best Practice: Enable quality assessment of open data. (July). Retrieved from <http://www.w3.org/2013/share-psi/bp/eqa-20160725/Latestversionhttp://www.w3.org/2013/share-psi/bp/eqa/Previousversionhttp://www.w3.org/2013/share-psi/bp/eqa-20160721/>
- Rafique, I., Lew, P., Abbasi, M. Q., & Li, Z. (2012). Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 6(5), 568–573. Retrieved from <http://www.waset.org/publications/9538>
- Ramos Simón, L. F., Arquero Avilés, R., Botezan, I., Cobo Serrano, S., Sala Jiménez, A., Sánchez Jiménez, R., & Valle Gastaminza, F. del. (2012). De la reutilización de información del sector público a los portales de datos abiertos en Europa. *BiD: Textos Universitaris de Biblioteconomia i Documentació*, 29. Retrieved from <http://bid.ub.edu/29/ramos2.htm%5Cnhttp://bid.ub.edu/29/pdf/ramos2.pdf>
- Reis, J. R., Viterbo, J., & Bernardini, F. (2018). A Rationale for Data Governance As an Approach to Tackle Recurrent Drawbacks in Open Data Portals. *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, 73:1--73:9. <https://doi.org/10.1145/3209281.3209354>

- Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, 37(3), 150–154. <https://doi.org/10.1016/j.ijinfomgt.2017.01.003>
- Smith, H. A., & McKeen, J. D. (2018). Developments in Practice XXX: Master Data Management: Salvation Or Snake Oil? *Communications of the Association for Information Systems*, 23. <https://doi.org/10.17705/1cais.02304>
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325–337. <https://doi.org/10.1016/j.giq.2016.02.001>