

Las mejores prácticas para compartir información del sector público

Proyecto Share-PSI 2.0

Septiembre 2017



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta

Este documento ha sido elaborado en el marco de la Iniciativa Aporta (datos.gob.es), desarrollada por el Ministerio de Asuntos Económicos y Transformación Digital a través de la Entidad Pública Empresarial Red.es

Aviso legal: Esta obra está sujeta a una licencia Atribución 4.0 de Creative Commons (CC BY 4.0). Está permitida su reproducción, distribución, comunicación pública y transformación para generar una obra derivada, sin ninguna restricción, siempre que se cite al titular de los derechos (Ministerio de Asuntos Económicos y Transformación Digital a través de la Entidad Pública Empresarial Red.es). La licencia completa se puede consultar en:
<https://creativecommons.org/licenses/by/4.0>



Share-PSI 2.0

[Share-PSI 2.0](#) es una red temática, cofinanciada por la Comisión Europea dentro del programa CIP (Competitiveness and Innovation Programme) y del marco de acciones de la [Agenda Digital Europa](#). El objetivo de dicha entidad es intercambiar experiencias sobre la implementación de las políticas de apertura de la información pública e impulsar el cumplimiento de la Directiva de Reutilización de Información del Sector Público (201/37/UE).

La red temática [Share-PSI 2.0](#) ha documentado una serie de buenas prácticas que sirven de orientación a todas las organizaciones públicas que cumplen con la Directiva de la Comisión Europea; al mismo tiempo ha ampliado y contextualizado las mejores prácticas de datos en la web del W3C ayudando a maximizar el potencial de la reutilización. Todas estas prácticas están enlazadas desde la [web del proyecto Share-PSI 2.0](#).

56 prácticas

El presente documento recoge las 56 best practices recopiladas en el proyecto Share PSI 2.0, revisándolas y presentándolas con las siguientes aportaciones:

- ✓ Nuevo formato de ficha que unifica las prácticas elaboradas por el grupo del Share-PSI con las provenientes del W3C.
- ✓ Disponibles en español.
- ✓ Clasificadas en 5 grupos en función de la fase del ciclo de vida de los datos en el que se esté trabajando: **estrategia, describir la información, publicar, acceso y medir y evaluar.**
- ✓ Enriquecimiento y actualización de los ejemplos de implantación y las referencias.
- ✓ Presentan un etiquetado por tema que trata y por los beneficios directos de adoptar dicha BP (impacto).
- ✓ Acceso a la versión original de la práctica en el web del proyecto Share-PSI mediante enlace en la esquina superior derecha.

Para cada una de las 56 de las best practices (BP) se ha elaborado una ficha con la siguiente información:

Audiencia de destino de la BP.

Título descriptivo de la BP	
 	
Breve resumen con los puntos principales de la BP.	        
Reto a solucionar	
Problema o reto que la BP pretende solucionar. Se incluirá la información de contexto.	
Implementación	Ejemplo de implementación
Posible forma de implementación de la BP.	Ejemplo de implementación donde se incluye código fuente, plantillas, etc.
Validación	
Descripción de los mecanismos para comprobar si la BP se ha aplicado correctamente	
Impacto	
      	
Referencias	
Enlaces a referencias relacionadas con la BP.	

Enlace al BP original

Temáticas principales de la BP

A continuación encontrarás las fichas de las BP organizadas en 5 grupos:

Estrategia

Describir la información

Publicar

Acceso

Medir y evaluar

ESTRATEGIA



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Publicador

Desarrollar un plan de publicación de datos abiertos (1/2)



Se deben identificar los conjuntos de datos que son aptos para la publicación, así como los requisitos de acceso de los interesados (tanto internos como externos). El plan de publicación de datos abiertos debe desarrollarse teniendo en cuenta las necesidades de las partes interesadas, así como los posibles beneficios, riesgos y costos de publicación de los conjuntos de datos.



Estrategia de publicación

Reto a solucionar

¿Cómo administrar la publicación de datos abiertos? Los organismos del sector público tienen un gran número de conjuntos de datos diferentes. Sin embargo, también operan bajo diversas restricciones, como el presupuesto, que les impiden hacer disponibles a la vez todos los conjuntos de datos potenciales para su reutilización.

Implementación

Desarrollar un plan de publicación de datos abiertos que equilibre los requisitos y las restricciones conocidas. Este plan de publicación debe contener todas las tareas necesarias para implementar la iniciativa Open Data y debe estar respaldado por una hoja de ruta para la publicación de los datasets seleccionados.

Se debe realizar un análisis de los conjuntos de datos disponibles, y se deben entender las necesidades y requerimientos tanto de las partes interesadas internas como externas y desarrollarse el plan de publicación de Open Data. La publicación de Open Data debe estar en línea con los objetivos estratégicos de la organización, con las políticas pertinentes, tales como la estrategia nacional de datos abiertos y debe satisfacer la demanda de los usuarios potenciales de los datos. Por lo tanto, los requisitos de las partes interesadas tanto internas como externas deben ser identificados y analizados. El plan de publicación Open Data debe contener todas las tareas necesarias para implementar la iniciativa Open Data. Con respecto a los conjuntos de datos que se prevé publicar como datos abiertos, debería elaborarse una hoja de ruta para su publicación. Se debería priorizar la publicación de los conjuntos de datos teniendo en cuenta los requisitos de las partes interesadas, así como los resultados del análisis de los conjuntos de datos, es decir, los beneficios y riesgos identificados y las estimaciones de coste y esfuerzo que esto conlleva. Las medidas de aplicación podrían resumirse de la siguiente manera:

- Obtener el apoyo de la alta dirección.
- Designar a una persona responsable del desarrollo del plan de publicación Open Data.
- Asegurar la cooperación de los departamentos involucrados en la iniciativa Open Data.
- Identificar los requerimientos de las partes interesadas tanto internas como externas.
- Desarrollar un plan de publicación Open Data que equilibre los requisitos y las restricciones conocidas para este fin.
- Desarrollar una hoja de ruta de liberación de datos abiertos basada en las prioridades del conjunto de datos.
- Establecimiento de un proceso de revisión para el plan de publicación Open Data.

Ejemplo de implementación

Algunos ejemplos de la implementación de esta BP son:

- Austria: [Open Government Data Implementation Model](#)
- España: [Basque Country PSI re-use assesment](#)
- Finlandia: [Helsinki Region Infoshare](#)

datos.gob.es facilita un listado [planes RISP de la administración pública](#), donde se incluyen las principales líneas de la política general de apertura y reutilización de la información del sector público (RISP) de cada organismo público implicado y "Evaluación en la adaptación al RD 1495/2011" respectivamente. Cada plan consta de los siguientes puntos y líneas de actuación:

- Antecedentes
- Identificación y selección de la información reutilizable
- Procedimiento de preparación de conjuntos de datos
- Condiciones de reutilización
- Diseño y gestión del espacio web dedicado
- Publicación y mantenimiento de los datos reutilizables
- Medidas de evaluación y mejora de la calidad en la reutilización y fomento del consumo de los mismos.



Publicador

Desarrollar un plan de publicación de datos abiertos (2/2)



Impacto



Reutilización de los datos



Procesamiento de los datos



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [Tendencias y buenas prácticas en la implementación de políticas de datos abiertos](#)
- 2012 datos.gob.es [ES]: [Guía de aplicación del Real Decreto 1495/2011](#)
- 2017 European data portal [EN]: [Extensive Study Recommendations for Open Data Portals: from setup to sustainability](#)
- 2016 Timisoara report [EN]: [Open data priorities and engagement — identifying data sets for publication: Report](#)
- 2013 Samos Workshop [EN]: [Open Data Publication Plan](#)
- Open Data Institute [EN]: [Engaging with reusers](#)
- Open Data Institute [EN]: [How to prioritise data to drive global development](#)



Publicador

Desarrollo e implementación de una estrategia interdepartamental



Esta buena práctica permite a las agencias comprender lo que se requiere, planificar en consecuencia y medir el progreso. Tener un plan estratégico es esencial para quienes toman decisiones al más alto nivel. Tanto la Carta de Datos Abiertos del G8, publicada en 2013, como el Shakespeare Review of Public Sector Information, enfatizaron la necesidad de un plan claro, visible y auditable para publicar datos tan pronto como sea posible.



Estrategia de publicación

Reto a solucionar

Diferentes organismos del sector público desarrollan e implementan su propia estrategia. Estas diferentes estrategias no están relacionadas, por lo que es difícil aplicar una planificación central y una evaluación.

Implementación

Para la implementación es necesario que exista una estrategia que coordine los esfuerzos de múltiples organismos.

Ejemplo de implementación

- Dar responsabilidad a un funcionario o departamento individual para desarrollar el plan.
- Convocar una reunión o una serie de reuniones entre las partes interesadas (publicadores de datos, usuarios de datos, etc.).
- Desarrollar el plan a través de un proceso iterativo antes de buscar respaldo de alto nivel.

Esta buena práctica se ha implementado en:

- Flanders, Belgium [The Flemish Innovation Projects](#)

Impacto



Estrategia de publicación



Interoperabilidad



Procesamiento de los datos

Referencias

- 2017 datos.gob.es [ES]: [Pautas metodológicas para la apertura de datos](#)
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2017 European data portal [EN]: [Extensive Study Recommendations for Open Data Portals: from setup to sustainability](#)
- 2016 European data portal [EN]: Analytical Report [Open Data and Cities](#)
- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos. Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar](#)
- 2017 European Data Portal [EN]: Practical Guide [Technical preparation and implementation](#)
- 2012 datos.gob.es [ES]: [Guía de aplicación del Real Decreto 1495/2011](#)
- 2013 Noël Van Herreweghe [EN]: [The Flemish Open Data Program](#)
- Open Data Institute guidance [EN]: [How to plan and budget an open data initiative](#)
- Open Data Institute white paper [EN]: [Open data in government, how to bring about change](#)
- Open Data Institute method report [EN]: [Supporting public sector open data leadership](#)
- Open Data Institute technical paper [EN]: [The Open Data Maturity Model and Pathway](#)



Publicador

Publicar un resumen de los datos gestionados



La comunidad de usuarios no siempre es totalmente consciente de la existencia de conjuntos de datos que una organización puede poner a su disposición como datos abiertos. Al menos no en su totalidad. Esto podría impedir a la comunidad proporcionar información útil sobre qué conjuntos de datos habría preferido para su publicación. Siguiendo esta práctica se mejoraría la eficiencia de la recopilación de información sobre los conjuntos de datos solicitados para su publicación.

Cuando se hace pública una visión general se proporciona suficiente información para que los usuarios (tanto del sector público como privado) prioricen los datos más interesantes para su publicación.

La transparencia sobre qué conjuntos de datos gestiona el sector público combinado con una simple clasificación del derecho legal de acceso a los datos equivale a agregar valor tanto al sector público como a los reutilizadores porque:

- El conocimiento de que estos datos existen tiene valor en sí mismo
- Permite a los usuarios aportar información sobre los datos que consideran prioritarios
- Se está proporcionando transparencia sobre la información que su agencia está administrando



Estrategia de publicación

Reto a solucionar

La cuestión es: ¿cómo asegurarse de que la comunidad es consciente de los conjuntos de datos que podrían estar disponibles como datos abiertos?

El sector público pregunta a la comunidad de usuarios qué datos les gustaría que se liberasen y priorizarán la publicación de datos basándose en esta retroalimentación. Sin embargo, la comunidad de usuarios a menudo no es consciente de lo que existe y por lo tanto no puede responder de manera significativa.

Implementación

Los requisitos técnicos son mínimos. Publicar una simple hoja de cálculo y un punto de contacto al que se pueda enviar información pueden ser suficientes. La visión general también se puede proporcionar dentro de un catálogo de datos abierto existente.

El impacto de la práctica puede mejorarse siguiendo la buena práctica "Categorise openness of data". Esto puede resolverse dentro de un catálogo de datos abierto existente utilizando la propiedad "access rights" en DCAT-AP.

Esto permite que los potenciales usuarios externos descubran los conjuntos de datos categorizados como legalmente restringidos (amarillo) o no públicos (rojo).

Ejemplo de implementación

Un ejemplo de la implementación de esta práctica es:

Helsinki Region Infoshare: [Helsinki Opens A Window to its Information Systems](#) (comunicado de prensa):

Impacto



Estrategia de publicación



Interoperabilidad



Acceso a los datos

Referencias

- 2015 Samos Workshop Story [EN]: [Traffic Light System For Data Sharing](#)
- 2014 Samos Workshop Talk [EN]: [Public Transport Data in the City of Gijon](#)



Publicador

Identificar lo que ya ha sido publicado



Con frecuencia, la información se publica en un proceso distribuido utilizando un sistema de gestión de contenido. Es posible que falte un inventario de activos de información ya publicados y puede ser difícil crearlo manualmente. Por lo tanto, las organizaciones podrían encontrar un reto en saber por dónde empezar y decidir qué recursos de información deben estar disponibles para su reutilización en formatos legibles por máquinas.



Calidad del dato



Estrategia de publicación

Reto a solucionar

¿Por dónde empezar cuando se decide qué recursos de información deben estar disponibles para su reutilización en formatos legibles por máquinas?

Identificar qué información debe estar disponible en formatos legibles por máquinas para su reutilización puede resultar un desafío debido a la falta de conocimiento de qué información se encuentra ya publicada y que la cantidad de esta información puede ser demasiado grande como para catalogarse manualmente.

Implementación

Una simple hoja de cálculo podría servir como un inventario de los activos de datos e información, pero dependiendo del volumen de información y los requisitos de la organización, se podrían implementar soluciones de catalogación como CKAN.

Se necesitan software y librerías que permitan la extracción de información, como Scrapy. Los metadatos recopilados utilizando este software se pueden utilizar como características para clasificar y agrupar los enlaces.

Ya sea creado manualmente o por medios automatizados, el inventario debe contener al menos metadatos básicos sobre los datos o recursos de información como el título, la ubicación, el formato actual y los términos de uso. Metadatos adicionales como la persona o unidad responsable, el formato de los datos o la frecuencia de actualización pueden ayudar a administrar, en el futuro, el proceso de publicación y podría ayudar a hacer estimaciones más precisas sobre el esfuerzo y los costos necesarios para publicar y mantener los conjuntos de datos abiertos.

Ejemplo de implementación

Un ejemplo de la implementación de esta práctica es:

Helsinki. [Region Inforshare Open Data Pipeline](#)

Impacto



Reutilización de los datos



Procesamiento de los datos



Estrategia de publicación

Referencias

- Open Knowledge [EN][ES]: [Open Data Handbook. ¿Cómo abrir los datos?](#)
- 2013 Timisoara Workshop Session [EN]: [Identifying what you already publish](#)
- 2013 Krems Workshop Session [EN]: [Extracting Structured Data from Unstructured Open Data](#)



Facilitar los datos de investigación



La ciencia abierta es un área enorme y compleja en la que el desempeño de la investigación tiene al sector público como uno de sus consumidores más importantes. En muchos casos, el sector público no puede obtener información actualizada y precisa sobre las investigaciones en curso debido a intereses comerciales o a una mala organización. Los beneficios de una plataforma nacional para la evaluación de la investigación hacen que los logros científicos sean más detectables y medibles, y por lo tanto puede afectar la innovación, la economía y la educación del país.



Estrategia de publicación



Negocio

Reto a solucionar

La ciencia abierta tienen algunas coincidencias con la información del sector público, ya que los resultados y estadísticas científicas son consumidos por el sector público para la toma de decisiones estratégicas, las evaluaciones, la educación y la gestión de la investigación. En muchos casos, el sector público no puede obtener información actualizada y precisa sobre las investigaciones en curso debido a intereses comerciales o a una mala organización. En la actualidad, la recopilación de los resultados de la investigación se encuentra fragmentada en la red. Esta falta de información puede conducir a una situación en la que no existe una opinión correcta sobre el nivel micro o macro de los resultados de la investigación nacional.

Implementación

Con la "Amsterdam Call for Action on Open Science", se revalorizan estos temas y se pone en marcha un esfuerzo significativo para encontrar y evaluar nuevas soluciones. Actualmente, muchos países establecen mandatos de acceso abierto para garantizar la visibilidad de las actividades de investigación a nivel nacional. Esto permite y anima a los investigadores a divulgar más información sobre su trabajo. Como paso siguiente, se pueden construir plataformas nacionales de monitoreo de la investigación para recolectar y organizar datos sobre la investigación. Esto se puede llevar a cabo en varios niveles:

- Compartiendo datos de registro sobre resultados de investigación y publicaciones
- Compartiendo textos completos de publicaciones
- Proporcionando una plataforma para archivar y compartir fácilmente los experimentos de investigación

Ejemplo de implementación

Un organismo del sector público debe asumir el papel de guía en el proceso. Se necesita un marco jurídico con varios elementos: declaración de objetivos, mandato de acceso abierto, ayuda para resolver cuestiones de derechos de autor a los editores y reglamentos para la provisión de datos. Se debe establecer una plataforma de IT con componentes centrales y distribuidos. Los requisitos de IT pueden variar en un amplio rango dependiendo de las tareas que se van a realizar. La logística de ciertos experimentos de investigación requiere de un gran espacio de almacenamiento y poder de cálculo. Además, los investigadores necesitan ser educados para entender por qué y cómo necesitan abrir su investigación.

Esta buena practica se ha implementado en :

- Suecia [Swepub](#)
- Hungría [MTMT](#)
- Finlandia [JUULI](#), [Etsin](#)
- Países Bajos [Narcis](#)

Impacto



Acceso a los datos



Interoperabilidad



Estrategia de publicación

Referencias

- Universidad Pontificia de Comillas [ES]: [OPENSOURCE: DATOS DE INVESTIGACIÓN EN ABIERTO](#)
- 2014 FECYT [ES]: [Recomendaciones para la implementación del artículo 37 Difusión en Acceso Abierto de la Ley de la Ciencia, la Tecnología y la Innovación](#)
- 2013 András Holl, András Micsik [EN]: [MTMT: The Hungarian Scientific Bibliography](#).
- 2013 Timișoara Workshop [EN]: [Role of Open Data in Research Institutions with International Significance \(notes\)](#)
- 2013 Robert Ulrich [EN]: [Making research data repositories discoverable](#)
- 2015 Robert Ulrich, Hans-Jürgen Goebelbecker, Frank Scholze, Michael Witt [EN]: [re3data.org - Making research data visible and discoverable](#)
- 2013 Krems Workshop bar camp session [EN]: [Open Science & Technology](#)



GOBIERNO DE ESPAÑA

MINISTERIO DE HACIENDA Y FUNCIÓN PÚBLICA

MINISTERIO DE ENERGÍA, TURISMO Y AGENDA DIGITAL

red.es

Iniciativa **aporta**



Publicador

Establecer un ecosistema de datos abiertos (1/2)



Las organizaciones están dando pasos agigantados hacia la apertura de datos públicos con el objetivo de ponerlos a disposición de todos, tanto para uso personal como para uso comercial. Sin embargo, la adopción de estos datos e información no ha sido lo que se esperaba. Poner los datos a disposición del público no es suficiente para que dichos datos sean útiles. Los ciudadanos no están interesados en los datos, sí lo están en los servicios que se están construyendo gracias a los datos e información disponibles. Por lo tanto, es necesario hacer más: establecer una red activa de datos abierta, un ecosistema para facilitar la adopción de datos e información para su reutilización.



Identificación



Estrategia de publicación



Documentación/
Catalogación



Comunidad

Reto a solucionar

El panorama de datos abiertos se compone de distintos elementos y usuarios con diferentes necesidades, desafíos, problemas, expectativas, roles y oportunidades previstas. Es necesaria una estrategia para coordinar a todos estos elementos internos y externos y a las partes interesadas. Sólo entonces podrán convencerse de compartir datos e información y ponerlos a disposición para su reutilización; sólo entonces todos los involucrados sabrán qué puede o debe estar disponible para realizar esfuerzos de innovación y nuevos productos y servicios previstos en la directiva PSI.

Implementación

Se necesita una red activa que facilite la interacción y comunicación entre todos los interesados y/o involucrados en en datos abiertos y la reutilización de información y datos, tanto internos como externos a la organización. Dentro de un ecosistema regional, nacional e internacional, el intercambio de información hará que aumente la visibilidad y, por lo tanto, la conciencia sobre la disponibilidad de datos e información para la reutilización y los resultados obtenidos al hacerlo. Las pequeñas empresas infomediarias podrían actuar como intermediarias entre todas las partes involucradas y con otros segmentos del mercado. Se pueden encontrar ejemplos en España donde más del 80% de las empresas infomediarias tienen más de 5 años y generan entre 4.500 y 5.500 puestos de trabajo, principalmente vinculados a las TIC: análisis, procesamiento y presentación de la información.

- Asegurarse de que todo está en su lugar para facilitar la implementación de datos abiertos (estrategia, contenido, legal y técnico)
- Estar dispuesto a adoptar un enfoque de abajo hacia arriba para implementar un programa de datos abiertos
- Lograr un ecosistema de datos abierto escuchando a todas las partes interesadas
- Dedicar las personas y recursos necesarios a la realización de este ecosistema
- Estar dispuestos a compartir datos dentro y entre las administraciones públicas
- Establecer los procedimientos de seguimiento necesarios para la continuidad

Ejemplo de implementación

Algunos ejemplos de implementación de esta BP son:

- Albania: [Utilization Cases of Open Data Albania](#), Julia Hoxha and Aranita Brahaj, Institute AIFB, Karlsruhe Institute of Technology, Karlsruhe, Germany and Albanian Institute of Science, Tirana, Albania. Presented at Share-PSI Samos Workshop Uses of Open Data Within Government for Innovation and Efficiency, July 2014.
- Finlandia: [Helsinki Loves Developers/](#)

Impacto



Reutilización de los datos



Confianza entre las partes



Publicador

Establecer un ecosistema de datos abiertos (2/2)



Referencias

- 2017 datos.gob.es [ES]: [Tendencias y buenas prácticas en la implementación de políticas de datos abiertos](#)
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 Open Data Monitor [EN]: [Open Data Topologies, Catalogues and Metadata Harmonization](#)
- 2014 datos.gob.es [ES]: [Informe de Tendencias e Iniciativas de Datos Abiertos](#)
- 2012 datos.gob.es [ES]: [Guía de aplicación del Real Decreto 1495/2011](#)
- 2013 M. D. Hernández, Ministerio de Hacienda y Administraciones Públicas [EN]: [Spanish Infomediary Sector Characteristics](#)
- 2013 Samos Workshop Talk [EN]: [Open Government Data Austria - Organisation, Procedures and Uptake](#)
- 2013 Noël Van Herreweghe [EN]: [The Flemish Open Data Program](#)
- 2013Lisbon Workshop Session [EN]: [Events, hackathons and challenge series - stimulating open data reuse](#)
- 2014 Lisbon Workshop Session [EN]: [Open Data Economy: from 'Wow' to 'How'](#)
- 2015 Szymon Lewandowski [EN]: [EU actions on Open Data – current policy and legal context](#)
- 2013 Krems Workshop Talk [EN]: [An Ongoing Open Dialog in an Open Data Ecosystem](#)
- 2013 Berlin Workshop Talk [EN]: [Government as a Developer - Challenges and Perils](#)
- Open Data Institute guidance [EN]: [Engaging with reusers](#)
- Open Data Institute method report [EN]: [Peer networks for open data leaders](#)
- Open Data Institute white paper [EN]: [How to improve agriculture, food and nutrition with open data](#)



Publicador

Modelos de negocio de datos abiertos y disciplinas de valor (1/2)



Por un lado, muchas organizaciones abiertas basadas en datos, específicamente a nivel de cadena de suministro, han diseñado, desarrollado e implementado con éxito su modelo de negocio, pero necesitan comprender el lado de la demanda (de las organizaciones menores) para dar un mejor servicio. Por otra parte, estas organizaciones menores tienen dificultades para desarrollar un modelo de negocio que les permita entender mejor esa demanda. Esta buena práctica ayuda a todos los tipos de organizaciones a superar los desafíos de desarrollar un modelo de negocio que sea eficaz para identificar lo que está sucediendo en la industria de datos abiertos y qué más se necesita hacer para apoyar y alimentar tanto la demanda como la oferta. Además, el desarrollo y la implementación de un modelo de negocio eficaz y eficiente puede conducir a una mayor satisfacción de los clientes y usuarios, haciendo emerger productos y servicios innovadores, generando ingresos y favoreciendo la supervivencia de la organización y, finalmente, conducir a maximizar el valor económico de los datos abiertos.



Negocio

Reto a solucionar

Todavía hay muchas organizaciones impulsadas por datos abiertos, especialmente a nivel medio y menores, que luchan por comprender cómo generar ingresos adaptándose a los cambios provocados por el continuo crecimiento de los datos abiertos y el "Big Data". Además, las organizaciones basadas en datos abiertos tienen dificultades para distinguir los diferentes modelos de negocio y entender cuál se adapta a su objetivo como organización.

Implementación

Para aprovechar el valor de los datos abiertos, maximizar los beneficios y permitir la creación de productos y servicios innovadores, las organizaciones basadas en datos deben desarrollar e implementar un modelo de negocio antes de iniciar su proyecto. Esto es necesario para asegurar que los productos y servicios generen propuestas de valor necesarias y satisfagan las necesidades de los clientes y usuarios y, finalmente, generen ingresos sustanciales. El Framework de 6 valores Open Data Business Model asegura que los gerentes tomen en cuenta todos los aspectos de un modelo de negocio eficaz y eficiente.

Ejemplo de implementación

Para poder comenzar a diseñar y desarrollar esta buena práctica, una organización necesita lo siguiente:

- Asociarse y tener experiencia necesaria (preferiblemente personas con conocimientos sobre datos abiertos y desarrollo de negocios).
- Paciencia al buscar en: el mercado, productos y servicios existentes de datos abiertos, colaboradores potenciales y competidores existentes.
- Definir un nicho de mercado.
- La disciplina de los valores de datos abiertos debe identificarse antes de desarrollar el modelo de negocio.

Las disciplinas de valor de datos abiertos ayudan a las organizaciones a concentrarse en ofrecer un valor superior al cliente. Los productos o servicios deben cumplir una o varias disciplinas de valor. Existen cuatro Disciplinas del Valor de Datos Abiertos:

1. Utilidad: adapta la propuesta de valor para apoyar directamente las necesidades de los consumidores de una forma u otra.
2. Mejora de Procesos: adapta la propuesta de valor para que coincida con las necesidades del cliente específicamente para mejorar los procesos.
3. Rendimiento: adapta la propuesta de valor para un mejor rendimiento.
4. Lealtad del cliente: adapta la propuesta de valor para fidelizar a los clientes.



Impacto



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [Estudio de casos de éxito y mejores prácticas en la reutilización de la información pública](#)
- 2017 datos.gob.es [ES]: [Estudio de Caracterización del Sector Infomediario 2016](#)
- 2015 European data portal [EN]: Extensive Study [Creating Value through Open Data](#)
- 2017 European data portal [EN]: Extensive Study [Re-using Open Data](#)
- 2015 European data portal [EN]: Analytical Report [Digital Transformation and Open Data](#)
- 2017 IDC [EN]: [European Data Market](#)
- 2013 Fatemeh Ahmadi Zeleti, Insight Centre for Data Analytics [EN]: [Open Data Business Model Generation](#)
- 2013 Charalampos Alexopoulos, Yannis Charalabidis [EN]: [Realising an Open Data Marketplace in Greece](#)
- 2013 Nicolas Hazard [EN]: [Business models for Linked Open Government Data: what lies beneath?](#)
- 2013 Tassilo Pellegrini, FH St. Pölten, Christian Dirschl & Katja Eck; Wolters Kluwer [EN]: [Linked Data Business Cube – Modelling Semantic Web business models](#)
- 2014 Fatemeh Ahmadi Zeleti, Adegboyega Ojo, Edward Curry [EN]: [Emerging Business Models for the Open Data Industry: Characterization and Analysis](#)
- Open Data Institute guidance [EN]: [How to make a business case for open data](#)
- Open Data Institute research [EN]: [Open data means business](#)
- Open Data Institute white paper [EN]: [Open enterprise: how three big businesses create value with open innovation](#)



Apoyar Start Ups de datos abiertos (1/2)



Un acelerador de negocios académico es una unidad organizativa dentro de una universidad que tiene como objetivo movilizar y apoyar a las personas para construir su propia empresa. La unidad es responsable de transferir una mentalidad emprendedora e innovadora a la próxima generación de expertos empresarios. También mantiene una colaboración con las estructuras privadas y públicas para la financiación y tutoría. Los datos abiertos pueden constituir una base muy útil para el espíritu empresarial, permitiendo el desarrollo de servicios con valor añadido para los ciudadanos y las pequeñas empresas. La subunidad de datos abiertos fomenta la colaboración entre las universidades (empresarios potenciales) y las organizaciones de financiación privadas y públicas (cámaras de comercio, municipios, inversores iniciales) y expertos del sector privado con el fin de fomentar las Start-ups de datos abiertos.



Reto a solucionar

Poner en marcha un negocio inspirado en RISP/datos abiertos requiere de una multitud de campos de conocimiento y mentoría.

Implementación

Las universidades están bien interconectadas, tienen el conocimiento necesario emprender iniciativas de datos abiertos (estadísticas, sistemas de visualización, programación) y suelen tener las relaciones necesarias con los profesionales y otros académicos para aportar las habilidades y conocimiento faltantes.

Además, pueden proporcionar:

- Recursos como instalaciones, materiales...
- Mentoring
- Soporte para las startups
- Potenciales empresarios

Se necesita de una nueva estructura en una universidad u otro centro educativo similar.

Pasos detallados para la creación de start-ups académicas:

- Hacer accesibles los datos abiertos a través de servicios web: proporcionar algunas herramientas iniciales, formación, grupos tecnológicos
- Exponer ejemplos de start-ups exitosas
- Organizar competiciones
- Conexión con el mundo exterior (contactar con periodistas)
- Prestar asesoramiento jurídico
- Los mentores también deben provenir de empresas con éxito
- Dejar constancia de todas las actividades o eventos realizados, involucrando a los estudiantes en el proceso
- Formar equipos con otras escuelas académicas (unir medios y técnicas)
- Proporcionar educación en comunicación intercultural, ya que es probable que las nuevas empresas ofrezcan servicios a escala internacional

Ejemplo de implementación

Algunos ejemplos de la implementación de esta práctica son:

- Grecia - [The University of the Aegean Startup Incubator](#)
- Grecia - [Gov4All platform](#): University of the Aegean/Microsoft Greece open data incubator
- Australia - [DataStart](#): A public-private partnership
- República Checa - [Fond Otakara Motejla](#)
- España - [Gijón OpenDataLab](#)

La iniciativa de datos abiertos del Gobierno de España [datos.gob.es](#) lanzó la primera convocatoria de [Premios Aporta](#) dirigidos a reconocer experiencias que se hayan puesto en funcionamiento durante los dos últimos años haciendo uso de datos abiertos públicos generados por las Administraciones Públicas o de datos abiertos públicos y datos generados por entidades privadas. El objetivo es fomentar la reutilización de conjuntos de datos abiertos públicos o privados en la búsqueda de soluciones innovadoras, con gran impacto y de utilidad para el sector público o para el sector privado.



Apoyar Start Ups de datos abiertos (2/2)



Impacto



Interoperabilidad

Referencias

- 2013 Lisbon Workshop Session [EN]: [Open Data Startups: Catalyzing open data demand for commercial usage](#)
- 2013 Krems Workshop Session [EN]: [University Business Accelerators on Open Data: Activities, Challenges and Best Practices](#)
- 2015 Yannis Charalabidis [EN]: [University Business Accelerators on Open Data: Activities, Challenges, and Best Practices](#)
- [Open Data Incubator Europe](#)
- Nicky van Oorschot, Bart van Leeuwen [EN]: [An Intelligent Fire Risk Monitor Based on Linked Open Data](#)



Fomentar el crowdsourcing alrededor del open data



Preparar la información del sector público para compartir puede ser lento, costoso y, a veces, difícil. La participación de la comunidad en esta tarea aumentará la calidad y la cantidad de los datos disponibles, así como incentivará y atraerá a los potenciales usuarios.



Reto a solucionar

Aumentar la calidad y la cantidad de datos legibles por máquinas dentro de un presupuesto restringido.

Implementación

La colaboración abierta distribuida (crowdsourcing) puede ser una forma eficaz de aumentar la calidad y la disponibilidad de datos legibles por máquinas, en particular para las instituciones del patrimonio cultural. Diversas técnicas innovadoras, como la gamificación, se pueden utilizar para aprovechar la habilidad y el entusiasmo de la comunidad en general. En un nivel práctico, los conjuntos de datos se pueden poner a disposición en plataformas como GitHub para que los usuarios puedan ofrecer correcciones (aceptar tales correcciones permanece bajo el control del propietario de los datos). Este es el enfoque adoptado por la Ciudad de Chicago. A nivel de político, la identificación de proyectos de crowdsourcing la margen las instituciones gubernamentales también puede ser un indicador de valiosos de los conjuntos de datos que deberían ser priorizados en la apertura de datos públicos ya que el nivel de participación de la comunidad es generalmente proporcional al nivel de interés en esos datos.

Pasos para implementar esta buena práctica:

- En primer lugar identificar la necesidad exacta y luego buscar grupos capaces de apoyar la solución de esa necesidad a través del crowdsourcing.
- Pensar en el crowdsourcing como otra herramienta para crear o mejorar conjuntos de datos y pensar en las fases del proyecto de recolección de datos y dónde el crowdsourcing podría encajar mejor.
- Involucrar a las partes interesadas que podrían beneficiarse de una fuente gratuita de ciertos conjuntos de datos y que les proporcionen fondos para sostener los esfuerzos de crowdsourcing.
- Las tareas deben ser pequeñas para poder ser completadas por voluntarios con tiempo limitado.
- Utilizar un enfoque de gamificación si es posible, es decir, motivación a través de aspectos lúdicos.
- Es posible utilizar crowdsourcing sin el conocimiento del usuario. El ejemplo más conocido de esto es el uso de CAPTCHAs para resolver la micro tarea de leer palabras que el software de reconocimiento óptico no puede y por ese método digitalizar textos difíciles de leer.

Ejemplo de implementación

Algunos ejemplos donde se ha implementado esta práctica son:

- Suecia: [Guiding principles for digital cultural heritage \(PDF\)](#)
- República Checa: [Czech Republic Společně otevřáme data](#)

Impacto



Acceso a los datos



Estrategia de publicación

Referencias

- 2016 Gwen Shaffer California State University Long Beach [EN]: [Crowdsourcing open data policies: Measuring impacts and improving outcomes](#)
- 2013, Malmö University Dimitris Paraschakis [EN]: [Crowdsourcing cultural heritage metadata through social media gaming](#)
- 2015 Krems Workshop Session [EN]: [Towards A Sustainable Austrian Data Market](#)



Soporte de alto nivel



Todas las acciones de datos abiertos necesitan apoyo a alto nivel para ser eficaces a la hora de superar esas objeciones y, no menos importante, asegurar los recursos y reorientar las prioridades. Cuando se hace correctamente, el patrocinador podría empezar por abogar las acciones por sí mismos.

Los organismos del sector público normalmente tienen sus propios procedimientos para obtener un apoyo de alto nivel, como el apoyo ministerial. A menudo, la manera más fácil de hacer público este apoyo es la inclusión de una introducción al documento por el funcionario superior. El soporte escrito en otros formatos (por ejemplo, comunicados de prensa, correo electrónico interno, medios sociales, página web, etc.) también podría ser beneficioso para las acciones de datos abiertos.



Estrategia de publicación



Negocio

Reto a solucionar

Las acciones de datos abiertos deben ser apoyadas por altos funcionarios que están capacitados para proporcionar autoridad de arriba hacia abajo cuando sea necesario.

Implementación

Para obtener un soporte de alto nivel para las acciones de datos abiertos se debe tener en cuenta los siguientes puntos:

- Hay que entender, ¿quién puede dar el apoyo que se necesita? ¿Qué tan amplia es la acción que se quiere llevar a cabo? ¿Quiénes son sus principales objetivos? Por ejemplo, las estrategias de los gobiernos nacionales normalmente necesitarán apoyo del ministerio, idealmente incluyendo al propio Jefe de Gobierno.
- Hay que ser simple y claro en el mensaje a la persona de quien se necesita la ayuda. Lo mejor es tener un proyecto listo para revisarse.
- Pensar en las formas de llegar a él/ella. Podría ser tan simple como un correo electrónico o podría incluir varias reuniones con funcionarios inferiores para convencerlos de esta necesidad.
- Pedir permiso para usar su nombre en diferentes medios. Lo más probable es que se tenga que coordinar estos aspectos con sus expertos en comunicación.

Ejemplo de implementación

Esta buena práctica tiene los siguientes ejemplos de implementación:

- En Flandes, desde 2011, los datos abiertos y los servicios compartidos han estado en el centro de la estrategia de eGovernment. Se creó el Open Data Programme para evidenciar que en una sociedad que funcione bien, sea transparente y democrática, los ciudadanos y las empresas deben tener acceso a los datos y la información del gobierno y compartir y reutilizar esa información libremente y con restricciones mínimas. El gobierno decidió un enfoque de arriba hacia abajo para implementar la estrategia de datos abiertos. Una nota conceptual fue redactada y firmada por todos los ministros.
- En Alemania, el éxito del gobierno abierto requiere de un amplio compromiso político en todos los niveles de gobierno debido a su importancia a nivel internacional y la necesaria transformación cultural en la formulación de políticas y en la administración. Será especialmente importante encontrar un patrocinador poderoso y otros partidarios de alto nivel para el proyecto que se identifiquen con el desarrollo del gobierno abierto.

Impacto



Confianza entre las partes



Estrategia de publicación

Referencias

- 2015 datos.gob.es [ES]: [Documentos de ayuda para un organismo RISIP](#)
- 2013 Noël Van Herreweghe [EN]: [The Flemish Open Data Program](#)

DESCRIBIR LA INFORMACIÓN



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Publicador

Proporcionar metadatos (1/2)



Proporcionar metadatos con contenido semántico es un requisito fundamental cuando se publican datos en la web para que tanto las personas como las aplicaciones informáticas puedan comprenderlos y tener información sobre aspectos importantes que les permitan hacer uso de ellos.



Documentación/
Catalogación

Reto a solucionar

La web es un espacio abierto de información sin contextualizar donde a menudo los publicadores y los consumidores de datos se desconocen entre sí. Esto significa que proporcionar metadatos es fundamental para poder localizar y utilizar estos datos. Es esencial proporcionar información que ayude a los usuarios y a las aplicaciones informáticas a comprender los datos, su estructura, así como otros aspectos importantes como la autoría, la licencia, la fecha de generación, etc.

Implementación

Opciones para proporcionar metadatos legibles por humanos:

- Proporcionar metadatos como parte de una página Web HTML
- Proporcionar metadatos como un archivo de texto independiente

Opciones para proporcionar metadatos legibles por máquinas:

- Los metadatos legibles por máquina pueden proporcionarse en un formato serializable como Turtle y JSON, o pueden incrustarse en la página HTML usando [HTML-RDFA] o [JSON-LD]. Si se publican por separado varios formatos, deben publicarse desde la misma URL mediante la negociación de contenido y estar disponibles en URI separadas que se distingan por la extensión del nombre de archivo.
- Cuando se definen metadatos legibles por máquina, se recomienda reutilizar términos estándar existentes y vocabularios populares. Por ejemplo, los términos de Dublin Core Metadata (DCMI) [DCTERMS] y el Catálogo de Datos Vocabulario [VOCAB-DCAT] se pueden utilizar para proporcionar metadatos descriptivos. Dichos vocabularios están diseñados para ser muy flexibles, por lo que a menudo es útil utilizar un perfil específico de un vocabulario como el DCAT-AP de la Comisión Europea.

Ejemplo de implementación

Para humanos

[Página de ejemplo](#) con una descripción legible por humanos de un conjunto de datos disponible

Para máquinas:

[Archivo de ejemplo](#) con una descripción legible por máquina en Turtle de un dataset disponible

Validación

Comprobar si están disponible los metadatos en formatos legible para humanos.

Comprobar si están disponible los metadatos en un formato legible para máquinas sin errores sintácticos

Impacto



Reutilización de los datos



Legibilidad de los datos



Publicador

Proporcionar metadatos (2/2)



Referencias

- 2017 datos.gob.es [ES]: [DCAT-AP y la Norma Técnica de Interoperabilidad de Reutilización de recursos de información \(NTI_RISP\)](#).
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 European data portal [EN]: Training module [Introduction to Metadata Management](#).
- 2017 European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2013 Universidad Técnica de Ingeniería Civil de Bucarest [EN]: [Free our maps](#)
- 2013 Share-PSI 2.0 [EN]: [Location Track](#)
- 2016 W3C [EN]: [Spatial Data on the Web Best Practices](#)



Proporcionar metadatos descriptivos (1/2)



Facilitar información descriptiva sobre los conjuntos de datos permite descubrir nuevos de manera automática en la web y, a su vez, permite a los usuarios entender más fácilmente la naturaleza de dichos conjuntos de datos y sus distribuciones



Reto a solucionar

Que las personas sean capaces de interpretar la naturaleza de los conjuntos de datos y sus distribuciones, y los agentes de software puedan descubrirlos automáticamente.

Implementación

Los metadatos descriptivos pueden incluir las siguientes características generales dentro de un conjunto de datos:

- El título y una descripción.
- Las palabras clave que lo describen.
- La fecha de publicación.
- La entidad responsable de ponerlo disponible.
- Un punto de contacto.
- La cobertura geográfica.
- El período temporal que cubre.
- La fecha de la última modificación.
- Los temas o categorías cubiertos.

Los metadatos descriptivos pueden incluir las siguientes características generales dentro de una distribución:

- El título y una descripción.
- La fecha de publicación.
- El formato.

La versión legible por máquinas de los metadatos descriptivos puede proporcionarse utilizando el vocabulario del catálogo de datos [VOCAB-DCAT].

Ejemplo de implementación

Legibilidad por humanos:

En la siguiente [página de ejemplo](#) se muestra una descripción legible por humanos de un conjunto de datos

Legibilidad por máquinas:

El siguiente ejemplo muestra cómo usar [VOCAB-DCAT] para proporcionar los metadatos de descubrimiento legibles por máquina para el dataset de paradas de autobús (stops-2015-05-05) con una distribución en formato CSV.

```

:stops-2015-05-05
  a dcat:Dataset ;
  dct:title "Bus stops of MyCity" ;
  dct:keyword "transport","mobility","bus" ;
  dct:issued "2015-05-05"^^xsd:date ;
  dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
  dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
  dct:spatial <http://sws.geonames.org/3399415> ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
  dcat:theme :mobility ;
  dcat:distribution :stops-2015-05-05.csv
.

:mobility
  a skos:Concept ;
  skos:inScheme :themes ;
  skos:prefLabel "Mobility"@en ;
  skos:prefLabel "Mobilidade"@pt
.

:themes
  a skos:ConceptScheme ;
  skos:prefLabel "A set of domains to classify documents"
.

:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8"

```



Publicador

Proporcionar metadatos descriptivos (2/2)



Ejemplo de implementación

En un ejemplo de una respuesta en formato XML proporcionada por la API REST Zaragoza, que provee conjuntos de datos que publicados por el Ayuntamiento de Zaragoza, se observa que se ofrecen metadatos como título, horario, teléfono, datación...:

```
<title>Museo del Foro de Caesaraugusta</title>
<description> El Foro es el centro neurálgico de la vida en una ciudad romana: es el principal lugar de reunión donde se desarrolla la vida político-administrativa, económica y religiosa...</description>
<estilo>romano</estilo>
<address>Plaza de la Seo, 2</address>
<horario> HORARIO SEMANA SANTA 2017:13,14 y 16 abril de 10 a 14.30h /horario>
<phone>976 72 12 21 - Reservas 976 72 60 75</phone>
<datacion>Siglo I a.C. - Siglo I d.C.</datacion>
<pois>Mercado de Augusto, Foro de Tiberio</pois>
```

Validación

Comprobar si los metadatos de los conjunto de datos incluyen las características generales del mismo en un formato legible por humanos.
Comprobar si los metadatos descriptivos están disponibles en un formato legible por máquinas válido.

Impacto



Reutilización de los datos



Legibilidad de los datos



Acceso a los datos

Referencias

- 2015 datos.gob.es [ES]: [Guía para publicar datos abiertos de manera rápida y sencilla \(con CKAN\)](#)
- 2017 datos.gob.es [ES]: [Conceptos básicos, beneficios del open data y barreras](#)
- 2017 datos.gob.es [ES]: [DCAT-AP y la Norma Técnica de Interoperabilidad de Reutilización de recursos de información \(NTI_RISP\)](#).
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2017 European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2014 European data portal [EN]: Training module [Introduction to Metadata Management](#)
- 2014 European data portal [EN]: Training Module [Promoting the Re-use of Open Data through ODIP](#)
- 2015 W3C [EN]: [R-MetadataAvailable](#)
- 2015 W3C [EN]: [R-MetadataMachineRead](#)
- 2015 W3C [EN]: [R-MetadataStandardized](#)



Publicador

Proporcionar metadatos estructurales (1/2)



Proporcionar información sobre la estructura interna de una distribución es esencial para los usuarios que deseen explorar o consultar un conjunto de datos. También ayuda a las personas a comprender el significado de los datos.



Calidad del dato



Documentación/ Catalogación

Reto a solucionar

Hacer que los seres humanos sean capaces de interpretar el esquema de un conjunto de datos y los agentes de software puedan procesar automáticamente las distribuciones.

Implementación

Normalmente, los metadatos estructurales legibles por humanos proporcionan las propiedades del esquema del dataset .

Los metadatos estructurales procesables por máquinas están disponibles según el formato específico de una distribución y pueden proporcionarse en documentos separados o incrustados en el documento. Para más detalles ver los siguientes enlaces:

- Datos tabulados. Véase [Modelo de datos tabulares y metadatos en la Web](#)
- JSON-LD. Véase [JSON-LD 1.0](#)
- XML. Véase [Esquema XML](#)
- Datos multidimensionales. Véase [Data Cube](#)

Ejemplo de implementación

Legible por humanos:

La siguiente [página de ejemplo](#) muestra metadatos estructurales legibles por humanos

Procesable por máquinas:

El siguiente ejemplo presenta los metadatos estructurales para una distribución del conjunto de datos de paradas de autobús:

```
{
  "@context": ["https://www.w3.org/ns/csvw", {
    "@language": "en"
  }],
  "url": "http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv",
  "dct:title": "CSV distribution of stops-2015-05-05 dataset",
  "dcat:keyword": ["bus", "stop", "mobility"],
  "dct:publisher": {
    "schema:name": "Transport Agency of MyCity",
    "schema:url": {
      "@id": "http://example.org"
    }
  },
  "tableSchema": {
    "columns": [{
      "name": "stop_id",
      "titles": "Identifier",
      "dct:description": "An identifier for the bus stop.",
      "datatype": "string",
      "required": true
    }, {
      "name": "stop_name",
      "titles": "Name",
      "dct:description": "The name of the bus stop.",
      "datatype": "string"
    } ] [ para más detalle, visitar https://www.w3.org/TR/dwbp/#StructuralMetadata ]
  }
}
```



Publicador

Proporcionar metadatos estructurales (2/2)



Validación

Comprobar si los metadatos estructurales se proporcionan en un formato legible por humanos.

Comprobar si los metadatos de la distribución incluyen información estructural sobre el conjunto de datos en un formato procesable por máquinas y sin errores de sintaxis.

Impacto



Reutilización de los datos



Legibilidad de los datos

Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2015 datos.gob.es [ES]: [Guía para publicar datos abiertos de manera rápida y sencilla \(con CKAN\)](#)
- 2017 datos.gob.es [ES]: [Conceptos básicos, beneficios del open data y barreras](#)
- 2014 European data portal [EN]: Training Module [Introduction to RDF & SPARQL](#)
- European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2015 W3C [EN]: [R-MetadataAvailable](#)



Publicador

Proporcionar información acerca de la procedencia de los datos



El origen de un conjunto de datos es una cualidad mediante la cual los consumidores juzgan su calidad. Comprender su origen e historia ayuda a determinar si es factible confiar en los datos y proporciona un importante contexto interpretativo.



Calidad del dato



Procedencia

Reto a solucionar

Hacer que las personas puedan conocer el origen y la historia del conjunto de datos y los agentes de software puedan procesar automáticamente esta información.

Implementación

La versión procesable máquinas de la información sobre el origen de los datos se puede proporcionar utilizando un metalenguaje, como la Ontología de Proveniencia (PROV-O) del W3C, recomendado para la descripción de esta información.

Ejemplo de implementación

En [esta página de ejemplo](#) se muestra información de procedencia legible por el usuario sobre un conjunto de datos de paradas de autobús:

En [este dataset](#) de la iniciativa de datos abiertos Open Data Euskadi se puede observar información de la procedencia en un formato legible por el usuario en la sección "Detalles".

Validación

Comprobar que los metadatos del conjunto de datos incluyen información de su origen en un formato legible por el usuario.
Comprobar si una aplicación informática puede procesar automáticamente la información de origen de un conjunto de datos.

Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2017 European data portal [EN]: [Towards an open government data ecosystem in Europe using common standards](#)
- 2015 W3C [EN]: [R-ProvAvailable](#)
- 2015 W3C [EN]: [R-MetadataAvailable](#)



Publicador

Proporcionar información sobre licencias (1/2)



La existencia de información acerca de licencias es esencial para que los consumidores de datos evalúen la usabilidad de los mismos. Los usuarios pueden usar la presencia o ausencia de información de licencias como un factor decisivo a la hora de incluir o no los datos para su reutilización.



Licencias

Reto a solucionar

Hacer que los humanos sean capaces de entender la información acerca de las licencias de datos donde describe posibles restricciones sobre el uso de un determinado dataset o distribución, y que los agentes software puedan detectar automáticamente la licencia de datos de una distribución.

Implementación

La información de licencia de datos puede estar disponible a través de un enlace, una copia incrustada o un acuerdo de licencia legible por humanos. También puede estar disponible para su procesamiento a través de un enlace o copia incrustada de un acuerdo de licencia procesable por máquinas.

Se puede utilizar uno de los siguientes vocabularios que incluyen propiedades para vincular a una licencia:

- Dublin Core [DCTERMS] (dct:license)
- Creative Commons [CCREL] (cc:license)
- schema.org [SCHEMA-ORG] (schema:license)
- XHTML [XHTML-VOCAB] (xhtml:license)

Ejemplo de implementación

Procesable por máquina:

La distribución CSV del dataset de paradas de autobús (stops-2015-05-05.csv) se publicará bajo la licencia Creative Commons Reconocimiento-Compartir igual 3.0 Unported. La propiedad dct:license se utiliza para incluir esta información como parte de los metadatos de distribución:

```
:stops-2015-05-05.csv
a dcat:Distribution ;
dct:title "CSV distribution of stops-2015-05-05 dataset" ;
dct:description "CSV distribution of the bus stops dataset of MyCity" ;
dcat:mediaType "text/csv;charset=UTF-8" ;
dct:license http://creativecommons.org/licenses/by-sa/3.0/
```

Legible por humanos:

La siguiente [página de ejemplo](#) muestra información acerca de la licencia que poseen los datos de la distribución

En [este dataset](#) de la iniciativa de datos abiertos de la Generalitat de Catalunya podemos encontrar información acerca de la licencia del uso de los datos expuestos en el.

Validación

Comprobar si los metadatos del conjunto de datos incluyen la información acerca de licencia de datos en un formato legible por humanos.
Comprobar si un agente informático puede detectar o descubrir automáticamente la licencia de datos del conjunto de datos.

Impacto



Reutilización de los datos



Interoperabilidad



Publicador

Proporcionar información sobre licencias (2/2)



Referencias

- 2015 datos.gob.es [ES]: [Guía para publicar datos abiertos de manera rápida y sencilla \(con CKAN\)](#)
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2017 datos.gob.es [ES]: [Buenas prácticas en la apertura de datos a lo largo del mundo](#)
- 2016 datos.gob.es [ES]: [Orientaciones sobre la protección de datos en la reutilización de la información del sector público](#)
- European data portal [EN]: Short overview of licences: [Introduction to licences for Open Data and explanation to the most often used licences](#)
- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos . Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar.](#)
- 2014 European data portal [EN]: Training Module [Data & metadata licensing](#)
- 2015 W3C [EN]: [R-LicenseAvailable](#)
- 2015 W3C [EN]: [R-MetadataMachineRead](#)
- 2015 W3C [EN]: [R-LicenseLiability](#)



Publicador

Proporcionar un indicador de versión (1/2)



La información acerca de la versión de un conjunto de datos permite que éste sea identificable de manera inequívoca. Esta singularidad de los conjuntos de datos puede ser utilizada por los consumidores para determinar si los datos han cambiado y cómo lo han hecho con el tiempo para determinar específicamente con qué versión de un conjunto de datos están trabajando. Llevar un control de las versiones de los datos permite dar a conocer a los consumidores si está disponible una versión más reciente de un conjunto de datos. El versionado explícito permite repetir la investigación, permite comparar y evita confusión. El uso de números de versión únicos siguiendo un estándar también puede ofrecer información a los consumidores sobre cuánto difieren las versiones.



Calidad del dato



Identificación



Estrategia de publicación

Reto a solucionar

Permitir que las personas y los agentes software sean capaces de determinar fácilmente con que versión de un conjunto de datos están trabajando.

Implementación

El mejor método para proporcionar información de control de versiones variará según el contexto. Sin embargo, hay algunas pautas básicas que se pueden seguir, por ejemplo:

- Incluir un número o fecha de versión único como parte de los metadatos del conjunto de datos.
- Utilizar un esquema de numeración coherente con un enfoque significativo para incrementar dígitos, como <https://www.w3.org/TR/dwbp/#bib-SchemaVer>
- Si los datos están disponibles a través de una API, el URI utilizado para solicitar la versión más reciente de los datos no debe cambiar a medida que cambian las versiones, pero debería ser posible solicitar una versión específica a través de la API.
- Usar Memento [RFC7089], o componentes de los mismos, para expresar el versionado temporal de un dataset y para acceder a la versión que estaba operativa en un datetime dado. El protocolo Memento se alinea estrechamente con el enfoque para asignar URIs a las versiones que se utilizan para las especificaciones W3C, que se describen a continuación.

El lenguaje de ontología web <https://www.w3.org/TR/dwbp/#bib-OWL2-QUICK-REFERENCE> y la ontología <https://www.w3.org/TR/dwbp/#bib-PAV> de Provenance, Authoring y versioning proporcionan varias propiedades de notación para la información de versión.

Ejemplo de implementación

Legibilidad por personas:
Esta [página de ejemplo](#) muestra información acerca del versionado de datos legible por humanos

Legibilidad por máquinas:
El siguiente ejemplo muestra los metadatos para paradas de bus con la inclusión de los metadatos para el control de versiones. Las propiedades OWL: versionInfo y pav: version se utilizan para denotar la versión del conjunto de datos:

```
:stops-2015-05-05
  a dcat:Dataset ;
  dct:title "Bus stops of MyCity" ;
  dcat:keyword "transport","mobility","bus" ;
  dct:issued "2015-05-05"^^xsd:date ;
  dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
  dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
  dct:spatial <http://sws.geonames.org/3399415> ;
  dct:publisher :transport-agency-mycity ;
  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dct:creator :adrian ;
  owl:versionInfo "1.0" ;
  pav:version "1.0"
```

Validación

Comprobar si los metadatos del conjunto de datos o de las distribuciones proporcionan un número o fecha de versión único en un formato legible por el usuario.
Comprobar si una aplicación informática puede detectar automáticamente y procesar el número de versión o la fecha de un conjunto de datos o distribución.



Publicador

Proporcionar un indicador de versión (2/2)



Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2015 W3C [EN]: [R-DataVersion](#)



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Publicador

Proporcionar el historial de versiones (1/2)



En la creación de aplicaciones que usan datos, puede resultar útil comprender la variabilidad de esos datos a lo largo del tiempo. La interpretación de los datos también se enriquece con la información de cómo cambia a lo largo del tiempo. Determinar cómo las diferentes versiones de un conjunto de datos se diferencian entre sí es suele ser muy laborioso a menos que se proporcione un resumen de las diferencias.



Calidad del dato



Identificación



Estrategia de publicación

Reto a solucionar

Se pretende que los seres humanos y los agentes software sean capaces de entender cómo los datasets cambian de una versión a otra y en que difieren dos versiones específicas.

Implementación

Proporcionar un listado de versiones publicadas y una descripción para cada versión que explique en que se diferencia cada una de la anterior. Una API puede exponer un historial de versiones con una única URL dedicada que recupere la última versión del historial completo.

Ejemplo de implementación

Legibilidad por humanos:
En la siguiente [página de ejemplo](#) se muestra información acerca del historial de versiones de datos legible por humanos

Legibilidad por máquinas:
Supongamos que se creó una nueva parada de bus y se publica un nuevo conjunto de datos (stops-2015-12-17) para mantener los datos actualizados. El nuevo conjunto de datos es una versión de stops-2015-05-05. Los metadatos legibles por máquina del nuevo conjunto de datos se muestran a continuación con la información del historial de versiones correspondiente:

```
:stops-2015-12-17
  a dcat:Dataset ;
  dct:title "Bus stops of MyCity" ;
  dcat:keyword "transport","mobility","bus" ;
  dct:issued "2015-12-17"^^xsd:date ;
  dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
  dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
  dct:spatial <http://sws.geonames.org/3399415> ;
  dct:publisher :transport-agency-mycity ;
  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
  dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ;
  dct:creator :adrian ;
  ...
  dct:isVersionOf :stops-2015-05-05 ;
  pav:previousVersion :stops-2015-05-05 ;
  rdfs:comment "The bus stops dataset was updated to reflect the creation of a new bus stop at 1115 Pearl Street."
;

owl:versionInfo "1.1" ;
pav:version "1.1"
.
```

Validación

Comprobar que esté disponible una lista de versiones publicadas así como un registro de cambios que describa con precisión en que cambia una versión respecto de la anterior



Publicador

Proporcionar el historial de versiones (1/2)



Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- W3C [EN]: [R-DataVersion](#)



Publicador

Asignar URIs a las versiones y las series de los conjuntos de datos



Al igual que en los documentos, muchos conjuntos de datos se pueden serializar o agrupar. Por ejemplo:

- Paradas de autobús en MyCity (que cambian con el tiempo);
- Una lista de funcionarios electos en MyCity
- La evolución de las versiones de un documento hasta su finalización.

En diferentes circunstancias, será apropiado referirse a la situación actual (el conjunto actual de paradas de autobús, los actuales funcionarios electos, etc.).

En otros, puede ser apropiado referirse a la situación tal como existía en un momento específico.



Calidad del dato



Estrategia de publicación

Reto a solucionar

Hacer que los seres humanos y los agentes software puedan referirse a versiones específicas de un conjunto de datos y a conceptos tales como a una serie de conjuntos de datos y la última versión de estos.

Implementación

El W3C proporciona un ejemplo de cómo hacerlo. Un URI persistente para de un documento es <https://www.w3.org/TR/2016/PR-dwbp-20161215/>. Ese identificador señala una versión del documento correspondiente al día de su publicación. El URI para la "última versión" de este documento es <https://www.w3.org/TR/dwbp/>, que es un identificador para una serie de documentos estrechamente relacionados que están sujetos a cambios en el tiempo. En el momento de la publicación, estos dos URIs identificarán al mismo documento. Sin embargo, cuando se publique la siguiente versión de este documento, se cambiará el URI de la "última versión" para que apunte a este, pero el URI con fecha permanece sin cambios.

Ejemplo de implementación

Supongamos que se crea una nueva parada de autobús. Para mantener las paradas-2015-05-05 actualizadas, se crea una nueva versión del conjunto de datos (stops-2015-12-17). Stops-2015-12-17 incluye todos los datos de stops-2015-05-05 más los datos sobre la nueva parada de autobús. Las dos versiones pueden ser identificadas por los siguientes URI:
[Http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05](http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05) es el URI de la primera versión del conjunto de datos
[Http://data.mycity.example.com/transport/dataset/bus/stops-2015-12-17](http://data.mycity.example.com/transport/dataset/bus/stops-2015-12-17) es la URI de la versión actualizada del conjunto de datos
[Http://data.mycity.example.com/transport/dataset/bus/stops](http://data.mycity.example.com/transport/dataset/bus/stops) siempre devuelve la última versión por lo que redirigirá a stops-2015-05-05 hasta el 17 de diciembre de 2015 cuando la configuración del servidor se actualice para que esa URL apunte stops-2015-12-17

Validación

Comprobar que cada versión de un conjunto de datos tiene su propio URI, y que además hay un URI que siempre devuelva la última versión.

Impacto



Reutilización de los datos



Legibilidad de los datos



Procesamiento de los datos

Referencias

- 2014 European data portal [EN]: Training Module [Design & Manage Persistent URIs](#)
- European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2015 W3C [EN]: [R-UniqueIdentifier](#)
- 2015 W3C [EN]: [R-Citable](#)



Publicador

Evaluar la cobertura del conjunto de datos (1/2)



Un fragmento de datos Web depende, por definición, del resto de un grafo global. Este contexto global influye en el significado de la descripción de los recursos encontrados en el conjunto de datos. Idealmente, la preservación de un determinado conjunto de datos implicaría preservar todo su contexto. Esto representaría toda la Web de Datos.

Al momento de archivar es necesario evaluar la vinculación entre el volcado de datos de los recursos ya preservados y los vocabularios utilizados. Los conjuntos de datos para los cuales muy pocos de los vocabularios utilizados y/o recursos apuntados están ya preservados en alguna parte deben ser marcados como en estado de riesgo.



Estrategia de publicación

Reto a solucionar

Posibilitar que los usuarios puedan hacer uso de datos archivados en el futuro.

Implementación

Compruebe si todos los recursos utilizados ya se han guardado en alguna parte o deben proporcionarse junto con el conjunto de datos que se considera para su conservación.

Ejemplo de implementación

Este conjunto de datos es adecuado para la preservación, ya que es auto-descriptivo y sólo hace uso de entidades externas cuya conservación es confiable.

```

<http://data.mycity.example.com/transport/route/bus/AB> a gdfs:Route;
  gdfs:color "ff0000" ;
  gdfs:shortname "10" ;
  gdfs:longName "Airport - Bullfrog" ;
  gdfs:agency <http://data.mycity.example.com/transport-agency/DTA> ;
  gdfs:routeType ex:three ;
  ex:usualVehicleType dbpedia:Routemaster ;
  foaf:isPrimaryTopicOf ex:Airport_Bullfrog
.

<http://data.mycity.example.com/transport/route/bus/BFC>
  a gdfs:Route;
  gdfs:color "ffff00";
  gdfs:shortname "20";
  gdfs:longName "Bullfrog - Furnace Creek Resort";
  gdfs:agency <http://data.mycity.example.com/transport-agency/DTA>;
  gdfs:routeType ex:three;
  ex:usualVehicleType dbpedia:Articulated_bus;
  foaf:isPrimaryTopicOf ex:Bullfrog_Furnace_Creek_Resort
.
...
# Custom vocabulary element
ex:usualVehicleType
  a rdf:Property ;
  rdfs:subPropertyOf gdfs:routeType ;
  rdfs:range gdfs:Bus
.

```

Validación

Es imposible determinar lo que estará disponible en, por ejemplo, 50 años. Sin embargo, se puede comprobar que un conjunto de datos archivado sólo depende de recursos externos ampliamente utilizados y vocabularios. Para ello hay que comprobar que las dependencias únicas o menos utilizadas se conservan como parte del archivo.



Publicador

Evaluar la cobertura del conjunto de datos (2/2)



Impacto



Confianza entre las partes



Reutilización de los datos

Referencias

- 2015 EC-Joinup [EN]: [dcat-ap](#)
- 2015 W3C [EN]: [R-VocabReference](#)



Enriquecer datos generando nuevos metadatos (1/2)



El enriquecimiento de los datos puede mejorar en gran medida su procesabilidad, particularmente para los datos no estructurados. En algunas circunstancias, se pueden rellenar valores incompletos y agregar nuevos atributos y medidas a partir de los datos ya existentes. Los conjuntos de datos también pueden enriquecerse reuniendo resultados adicionales de la misma manera que los datos originales o combinando los datos originales con otros conjuntos de datos. Publicar conjuntos de datos más completos puede mejorar la confianza entre las partes, si se hace correctamente. Derivar valores adicionales que son de utilidad general ahorra tiempo a los usuarios y estimula más tipos de reutilización. Hay muchas técnicas que pueden usarse para enriquecer datos, haciendo que el conjunto de datos sea un activo aún más valioso.



Reto a solucionar

Se pretende que se consigan rellenar los valores faltantes de los conjuntos de datos. Además, añadiendo medidas o atributos relevantes, se consigue mejorar la estructura y la utilidad de los datos, pero sólo si esta adición no distorsiona los resultados analíticos, el significado o la potencia estadística de estos.

Implementación

Las técnicas para el enriquecimiento de datos son complejas y van mucho más allá del alcance de este documento, por lo que sólo se resaltarán ciertas de sus posibilidades.

Mecanismos de "learning machine" puede aplicarse fácilmente al enriquecimiento de los datos. Estos métodos incluyen aquellos enfocados en categorización de datos, desambiguación, reconocimiento de entidad, tipificación, entre otros. Los nuevos valores de datos se pueden derivar de realizar un cálculo matemático entre las columnas existentes. Otros ejemplos incluyen la inspección visual como medio para identificar características en datos espaciales y referencias cruzadas a bases de datos externas para proporcionar información demográfica. Por último, la generación de nuevos datos puede estar orientada a la demanda, en la que los valores perdidos se calculan o determinan por medios directos.

Los valores generados por técnicas basadas en la inferencia deberían ser etiquetados como tales, y debería ser posible recuperar cualquier valor original reemplazado por enriquecimiento. Siempre que la licencia lo permita, el código utilizado para enriquecer los datos debería estar disponible junto con el conjunto de datos. Compartir este código es particularmente importante para los datos científicos.

La priorización de las actividades de enriquecimiento así como en el esfuerzo requerido deben basarse en el valor que esto conlleva para el consumidor de datos. El valor para el consumidor puede medirse mediante la medición de la demanda (por ejemplo, mediante encuestas o seguimiento de solicitudes).

Ejemplo de implementación

1. La agencia de transporte MyCity tiene direcciones de calles para cada una de sus paradas de tránsito. Quiere hacer más fácil para los consumidores de sus datos combinar los datos con mapas, por lo que agrega información de latitud y longitud para cada parada utilizando una base de datos geográfica.
2. La agencia de tránsito tiene una gran número de correos electrónicos de los usuarios. Algunos de los correos son de cortesía, otros son quejas, y algunos son solicitudes de información. La agencia realiza una combinación de análisis de sentimientos y categorización para extraer metadatos para cada uno de los mensajes, como el medio de transporte usado, la ruta usada o la actitud del conductor, para crear un conjunto de datos semiestructurado.

Validación

Comprobar que no faltan valores en el conjunto de datos o campos adicionales que podrían proporcionarse fácilmente y que probablemente sean necesarios para otros usuarios. Comprobar que cualquier dato agregado por técnicas de enriquecimiento se identifique como tal y que los datos reemplazados estén todavía disponibles.

Impacto



Reutilización de los datos



Legibilidad de los datos



Procesamiento de los datos

Referencias

- 2014 European data portal [EN]: Training module [Introduction to Metadata Management](#).
- 2015 W3C [EN]: [R-DataEnrichment](#)
- 2015 W3C [EN]: [R-FormatMachineRead](#)
- 2015 W3C [EN]: [R-ProvAvailable](#)



Citar la publicación original



Los datos sólo son útiles cuando son dignos de confianza. Identificar la fuente de los datos es un indicador importante de confiabilidad por dos razones: en primer lugar, el usuario puede juzgar la confiabilidad de los datos dada la reputación de la fuente, y segundo, citando la fuente sugiere que usted mismo es digno de confianza como un reeditor. Además de informar al usuario final, citar ayuda a los publicadores a acreditar su trabajo. Los publicadores que hacen que los datos estén disponibles en la Web merecen ser reconocidos y es más probable que continúen compartiendo datos si se reconoce su trabajo. Citar la fuente también mantiene la procedencia y ayuda a otros a trabajar con los datos.



Procedencia



Estrategia de publicación



Documentación/Catalogación

Reto a solucionar

El objetivo es que los usuarios finales sean capaces de evaluar la confiabilidad de los datos y que los esfuerzos de los publicadores sean reconocidos.

Implementación

Presentar la cita a la fuente original en una interfaz de usuario proporcionando texto bibliográfico y un enlace a dicha fuente.

Ejemplo de implementación

Fuente: MyCity Transport Agency. "Bus timetable of MyCity" (series 1.2). MyCity. May 5, 2015. Disponible en: <http://data.mycity.example.com/transport/dataset/bus/stops>.

En el apartado [Cómo utilizar los datos – Fórmulas para la citación de la fuente](#) de la página web del proyecto de datos abiertos Generalitat de Catalunya se especifica la fórmula exacta en que se tiene que citar en los datos por parte de las empresas o usuarios reutilizadores:

Fuente: Generalitat de Cataluña. [nombre del organismo o ente autónomo]
Si se incluye esta cita en formato HTML, puede utilizar el marcado siguiente, o similar:
Fuente de los datos: Generalitat de Cataluña.

Validación

Comprobar que la fuente original de los datos reutilizados se cite en los metadatos proporcionados. Comprobar que se trata una cita legible por el usuario y que sea fácilmente visible en cualquier interfaz de usuario

Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2015 W3C [EN]: [R-Citable](#)
- 2015 W3C [EN]: [R-ProvAvailable](#)
- 2015 W3C [EN]: [R-MetadataAvailable](#)
- 2015 W3C [EN]: [R-TrackDataUsage](#)



Publicador

Categorizar la apertura de los datos



El establecimiento de un sistema simple para categorizar la apertura de los datos facilita a las organizaciones del sector público el determinar con quién compartir datos.



Identificación



Estrategia de publicación

Reto a solucionar

La dificultad radica en cómo asegurarse que datos se comparten sólo con usuarios autorizados e identificar los conjuntos de datos que pueden abrirse fácilmente para todos. Las organizaciones del sector público a menudo sólo consideran los datos que pueden estar abiertos para todos o se preocupan por los datasets que difícilmente pueden ser abiertos. Sin embargo, también puede ser útil publicar datos que se pueden compartir bajo ciertas restricciones.

Implementación

Las descripciones tanto de datos abiertos como de datos con restricciones legales se pueden publicar con una indicación de con quién se puede compartir dichos datos. La propiedad "access right" de DCAT-APs se puede utilizar para este fin, combinado con un sistema de código de colores para los usuarios finales:

- :public (verde) para los datos que se pueden poner a disposición sin restricciones
- :restricted (amarillo) para los datos que no están abiertos para todos y tiene algunas restricciones
- :non public (rojo) para los datos que son sensibles y sólo pueden estar disponibles bajo estrictas condiciones.

Un código de color no es la única manera de denotar la categoría de apertura de datos. Otros esquemas con un significado similar podrían utilizarse como bien.

Ejemplo de implementación

Un ejemplo de la implementación de esta práctica es:

- Noruega: [Norway Difi Traffic Light System](#)

Impacto



Reutilización de los datos



Acceso a los datos

Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2015 Samos Workshop Story [EN]: [Traffic Light System For Data Sharing](#)
- Open Data Institute guidance [EN]: [What makes data open?](#)
- Open Data Institute guidance [EN]: [What are the impacts of non-open licences?](#)



Publicador

Elegir el nivel de formalización correcto (1/2)



La semántica formal ayuda a establecer especificaciones precisas que transmiten un significado detallado y el uso de un vocabulario complejo (ontología) puede servir como base para tareas tales como el razonamiento automatizado. Por otro lado, los vocabularios complejos requieren de más esfuerzo para la producción y comprensión de los datos, lo que podría obstaculizar su reutilización, comparación y vinculación con conjuntos de datos que los utilizan. Si los datos son suficientemente ricos para soportar preguntas detalladas de investigación (el hecho de que A, B y C son verdaderas, y que D no es verdad, conduce a la conclusión E), entonces algo como un perfil OWL sería claramente apropiado. Sin embargo, no hay nada complicado en una lista de paradas de autobús.

Elegir un vocabulario muy sencillo siempre es atractivo, pero existe el peligro de impulsar la simplicidad que podría llevar al publicador a omitir algunos datos que proporcionan información importante, como la ubicación geográfica de las paradas de autobús que impedirían mostrarlas en un mapa. Por lo tanto, tiene que existir un cierto equilibrio, recordando que la meta no es simplemente compartir sus datos, sino también que otros los reutilicen.



Calidad del dato



Documentación/
Catalogación

Reto a solucionar

Hacer que los casos de aplicación más probables sean compatibles con una complejidad no mayor de la necesaria.

Implementación

Se puede encontrar cierto vocabulario que se podría usar, pero este presenta una restricción semántica que hace que sea difícil utilizarlo, como puede ser una restricción de dominio o rango que no se aplica al caso dado. En ese escenario, a menudo vale la pena ponerse en contacto con los editor del vocabulario y tratar con ellos al respecto. Pueden ser capaces de levantar esa restricción y proporcionar más orientación sobre cómo utilizar el vocabulario de forma más amplia.

W3C opera una lista de correo en public-vocabs@w3.org donde se pueden discutir temas relacionados con el uso y desarrollo del vocabulario.

Si está creando un vocabulario propio, mantenga las restricciones semánticas al mínimo para aumentar la posibilidad de reutilización por parte de otros. Como ejemplo, los diseñadores de la ontología SKOS (ampliamente utilizada) han minimizado su compromiso ontológico cuestionando todos los axiomas formales sugeridos para sus clases y propiedades.

Otro ejemplo de este "design for wide use" se puede ver en schema.org. Lanzado en junio de 2011, schema.org fue adoptado masivamente en muy poco tiempo en parte debido a su enfoque informativo más que normativo para definir los tipos de objetos con los que se pueden usar las propiedades.

Ejemplo de implementación

Se codifican los datos de la parada de autobús usando GTFS (<https://www.w3.org/TR/dwbp/#bib-GTFS>) porque:

- Presenta un uso generalizado;
- Ofrece un nivel de detalle que coincide con los necesarios para sus datos
- Una motivación para la publicación de datos de parada de autobús es para apoyar el desarrollo de aplicaciones para ayudar a los usuarios de autobuses y GTFS está diseñado para ese propósito.

Validación

Este tema es casi siempre una cuestión de juicio subjetivo sin prueba objetiva. Como guía general se deberían formular las siguientes cuestiones:

- ¿Se utilizan vocabularios comunes como Dublin Core y schema.org?
- ¿Los datos sencillos son declarados de manera simple y pueden ser recuperados fácilmente?



Publicador

Elegir el nivel de formalización correcto (2/2)



Impacto



Reutilización de los datos



Legibilidad de los datos



Procesamiento de los datos

Referencias

- European data portal [EN]: [Controlled vocabularies: Learn what controlled vocabularies are and how to use them](#)
- [Linked Open Vocabularies](#) [EN]
- 2015 W3C [EN]: [R-VocabReference](#)
- 2015 W3C [EN]: [R-QualityComparable](#)

PUBLICAR



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Publicador

Usar URIs persistentes como identificadores de conjuntos de datos (1/2)



La adopción de un sistema de identificación común permite la identificación y la comparación de los datos básicos por cualquiera de las partes interesadas de manera fiable, lo que se considera una condición previa esencial para la correcta gestión y reutilización de los datos. Dado que publicadores y consumidores pueden construir URIs en su código, es importante que esos URIs sean persistentes y que hagan referencia al mismo recurso a lo largo del tiempo sin la necesidad de intervención humana.



Identificación

Reto a solucionar

Hacer que los conjuntos de datos y/o la información se encuentren accesibles a lo largo del tiempo, independientemente del estado, la disponibilidad o el formato de los datos.

Implementación

Para ser persistentes, los URI deben diseñarse como tales. Véase, por ejemplo, el Estudio de la Comisión Europea sobre URIs persistentes [PURI], que a su vez enlaza con muchos otros recursos.

Cuando a un editor de datos no le es posible gestionar la persistencia de URIs, un enfoque alternativo es utilizar un servicio de redirección como Identificadores permanentes para la web o purl.org. Estos proporcionan URIs persistentes que pueden ser redirigidos según sea necesario para que la ubicación final pueda ser efímera. El software detrás de tales servicios está disponible de forma libre para que pueda ser instalado y administrado localmente si es necesario.

Los identificadores de objetos digitales (DOI) ofrecen una alternativa similar. Estos identificadores se definen independientemente de cualquier tecnología Web, pero pueden añadirse a un "URI stub". Las DOI son una parte importante de la infraestructura digital para los datos y bibliotecas de investigación.

Ejemplo de implementación

La URI <http://data.mycity.example.com/transport/dataset/bus/stops> tiene varias características que contemplan la persistencia:

- Todos los nombres están sujetos a cambios con el tiempo, pero al elegir un nombre de dominio, es razonable que se asuma que MyCity continuará existiendo y que continuará teniendo un gobierno. Por lo tanto, mientras que casos como Yugoslavia demuestran que incluso los nombres de los países cambian y los dominios de nivel superior desaparecen (como .yu), un nombre de dominio basado en el nombre de la ciudad es tan persistente como cualquier nombre de dominio puede ser.
 - Al poner datos en el subdominio data.mycity.example.com, se está creando un dominio específico que se puede administrar independientemente de cualquier departamento en particular.
 - No es seguro asumir que un departamento específico persistirá. Las autoridades de MyCity podrían decidir que la Agencia de Transportes deben fusionarse con otra para crear la Agencia de Transporte y Medio Ambiente. Por lo tanto, es correcto no incluir el nombre de la Agencia de Transportes en la URI, sino incluir la tarea de la que provienen los datos, en este caso la de proporcionar transporte público.
 - El segmento de ruta de datos /dataset es una indicación de que el URI identifica un dataset, en lugar de una ruta específica.
 - Del mismo modo, el segmento de ruta de /bus nos lleva más hacia el conjunto de datos específicos.
 - Finalmente, /stops nos lleva al conjunto de datos relativo a las paradas de autobús en MyCity.
- En los términos DCAT, este sería el identificador del conjunto de datos. Es probable que las distribuciones específicas del conjunto de datos se identifiquen agregando la extensión de archivo relevante al URI, como <http://data.mycity.example.com/transport/dataset/bus/stops.csv>, <http://data.mycity.example.com/transport/dataset/bus/stops.json>, <http://data.mycity.example.com/transport/dataset/bus/stops.ttl>.



Publicador

Usar URIs persistentes como identificadores de conjuntos de datos (2/2)



Ejemplo de implementación

En la sección de [Buenas prácticas](#) del portal de datos abiertos de la Junta de Castilla y León, en el desarrollo del mismo se marcó como objetivo disponer de un esquema de URIs estable, persistente y extensible, de forma que se proporcione la base mínima necesaria para habilitar el mecanismo de Reutilización de la Información Pública en el Gobierno de Castilla y León.

Algunas de las características deseadas, recogidas en las buenas prácticas para la utilización de URIs en la Web Semántica y que se han seguido a la hora de diseñar el esquema de URIs son las siguientes:

- Utilizar siempre el protocolo HTTP para garantizar que los URIs pueden ser resueltas.
- Utilizar una estructura de URIs consistente, extensible y persistente.
- Crear URIs que, en la medida de lo posible, sean comprensibles y significativas.
- No exponer información sobre la implementación técnica de los URIs (por ejemplo no utilizar las extensiones correspondientes a tecnologías como .php, .jsp, etc).

Validación

Comprobar que cada conjunto de datos se identifica mediante un URI que ha sido diseñado para la persistencia. Lo ideal sería que el sitio Web pertinente incluya una descripción del esquema de diseño y un compromiso creíble de persistencia si el editor ya no puede mantener el espacio URI por sí mismo.

Impacto



Reutilización de los datos



Interoperabilidad



Acceso a los datos

Referencias

- 2017 datos.gob.es [ES]: [DCAT-AP y la Norma Técnica de Interoperabilidad de Reutilización de recursos de información \(NTI_RISP\)](#).
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 European data portal [EN]: Training Module [Design & Manage Persistent URIs](#)
- 2017 European data portal [EN]: [Towards an open government data ecosystem in Europe using common standards](#)
- 2017 European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2015 W3C [EN]: [R-UniqueIdentifier](#)
- 2015 W3C [EN]: [R-Citable](#)



Publicador

Usar URIs persistentes como identificadores dentro de conjuntos de datos (1/2)



Los datos se vuelven más valiosos si se refieren o enlazan a los datos de otras entidades sobre el mismo aspecto, el mismo lugar, el mismo concepto, el mismo evento, la misma persona, etc. Eso significa usar los mismos identificadores en los conjuntos de datos y asegurarse de que sus identificadores puedan ser usados como referencia por otros conjuntos de datos. Cuando esos identificadores son URIs HTTP pueden ser buscados y así, más datos pueden ser descubiertos.

Bajo estas ideas se fundamenta la razón de ser de las 5 estrellas del Linked Data donde un punto de datos enlaza a otro, y de la hipermedia donde los enlaces pueden apuntar a datos adicionales o a servicios que pueden actuar o relacionarse con los datos de alguna manera.



Identificación



Estrategia de publicación

Reto a solucionar

Se pretende que los datos estén relacionados a través de la Web creando un espacio de información global accesible tanto para humanos como para máquinas.

Implementación

Los desarrolladores saben que a menudo el problema que están tratando de resolver ya ha sido resuelto por otras personas. De la misma manera, si se está buscando un conjunto de identificadores para cosas obvias como países, monedas, temas, especies, proteínas, ciudades y regiones- alguien ya lo habrá hecho. Los pasos para descubrir los vocabularios existentes pueden ser:

- Asegurarse de que los conjuntos de URI que utiliza sean publicados por un grupo u organización de confianza
- Asegurarse de que los conjuntos de URI tengan URIs persistentes.

Si no se encuentra un conjunto existente de identificadores que satisfagan sus necesidades, entonces se creará uno propio, siguiendo los patrones de persistencia de URI para que otros agreguen valor a sus datos enlazándolos

Ejemplo de implementación

La siguiente URI dada como ejemplo: <http://data.mycity.example.com/transport/dataset/bus/stops>, identifica un conjunto de datos. Gran parte del URI puede reutilizarse para identificar paradas de autobús, rutas y el tipo de autobús utilizado en un servicio dado. Por ejemplo, un URI persistente adecuado para la ruta 'Aeropuerto-Bullfrog' sería:

<http://data.mycity.example.com/transport/route/bus/id/AB>

Este URI tiene la misma estructura inicial que para el conjunto de datos, pero en lugar de /dataset ahora se incluye el camino de segmento /route para que los usuarios puedan ver que lo que se identifica es una ruta de autobús. El segmento /id indica que el URI identifica algo que no es un recurso de información, es decir, algo que no se puede recuperar a través de Internet, y /AB es el identificador local para la ruta de bus real. Esto es consistente con el consejo de la guía de implementación de GS1 SmartSearch que dice que cuando se utilizan identificadores estándar para un producto, ubicación, etc., se recomienda que el URI incluya el tipo de identificador que se está utilizando. Por ejemplo, si se está utilizando un GTIN para identificar un producto, el URI debe tener el formato siguiente:

<http://data.myproduct.example.com/gtin/05011476100885>. Hace referencia a URIs para recursos que no son de información y debe dar como resultado una redirección HTTP 303 a una URL similar como <http://data.mycity.example.com/transport/route/bus/doc/AB> que describe, es decir, proporciona información acerca de AB ruta de autobús (tenga en cuenta la sustitución de /doc for /id).

Al ofrecer este consejo, se reconoce que los URIs pueden ser largos. En un conjunto de datos de tamaño moderado, el almacenamiento de cada URI es probable que se haga repetitivo e inútil. En su lugar, defina localmente identificadores únicos para cada elemento (como AB en este ejemplo) y proporcione datos que les permitan convertirlos a URIs globalmente únicos mediante la implementación de algún programa. El Metadata Vocabulary for Tabular Data [Tabular-Metadata] proporciona mecanismos para hacer esto dentro de datos tabulares como archivos CSV, en particular utilizando las propiedades de la plantilla URI, como la propiedad about URL.



Publicador

Usar URIs persistentes como identificadores dentro de conjuntos de datos (2/2)



Validación

Comprobar que dentro del conjunto de datos, las referencias sus elementos que no cambian o que cambian lentamente, como países, regiones, organizaciones y personas, son referidas por URIs o por identificadores cortos que se pueden anexas a una URI mas completa.

Impacto



Interoperabilidad



Acceso a los datos



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [DCAT-AP y la Norma Técnica de Interoperabilidad de Reutilización de recursos de información \(NTI RISP\)](#).
- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 European data portal [EN]: Training Module [Design & Manage Persistent URIs](#)
- 2017 European data portal [EN]: [Towards an open government data ecosystem in Europe using common standards](#)
- European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2015 W3C [EN]: [R-UniqueIdentifier](#)



Publicador

Conservar los identificadores



Localizar y consultar URIs es la forma principal de obtener datos en la Web. Si un URI da como respuesta un 404 (Recurso no encontrado), el usuario no sabrá si la falta de disponibilidad es permanente o temporal, planificada o accidental. Si el editor o un tercero ha archivado los datos, es mucho menos probable que se encuentre la copia archivada si el URI original está roto.



Identificación

Reto a solucionar

El objetivo es hacer que el URI de un recurso siempre lleve al recurso o redirija a la información sobre él.

Implementación

Hay dos escenarios a considerar:

1. El recurso se ha eliminado completamente y ya no está disponible a través de ninguna ruta
2. El recurso se ha archivado y sólo está disponible a través de una solicitud al archivo.

En el primero de estos casos, el servidor debe configurarse para responder con un código de respuesta HTTP de 410 (Gone).

En el segundo caso, donde los datos han sido archivados, es más apropiado redirigir las solicitudes a una página Web que da información sobre el archivo que contiene los datos y sobre cómo un usuario potencial puede acceder a él.

En ambos casos, el URI original continúa identificando el recurso y conduce a información útil, aunque los datos ya no están disponibles directamente.

Ejemplo de implementación

Se decide archivar las versiones de los datos de las paradas de autobús que han sido reemplazadas por versiones más recientes. El servidor se configura de tal manera que las solicitudes para el conjunto de datos antiguos se redirigen, utilizando el código HTTP 303, a una página Web que incluye la siguiente notificación:

Los datos que ha solicitado han sido archivados. Esto se encuentra en consonancia con la política de MyCity de archivar datos que fueron reemplazados hace más de 12 meses. Sin embargo, se puede solicitar una copia en cualquier momento a través de la página de contacto.

Validación

Comprobar que el URI de un conjunto de datos que ya no está disponible devuelve información sobre su estado actual y disponibilidad, utilizando un código de respuesta 410 o 303 según corresponda

Impacto



Reutilización de los datos



Confianza entre las partes



Estrategia de publicación

Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 European data portal [EN]: Training Module [Design & Manage Persistent URIs](#)
- 2015 W3C [EN]: [R-AccessLevel](#)
- 2015 W3C [EN]: [R-PersistentIdentification](#)



Crterios para conjuntos de datos (1/2)



Esta buena práctica establece una serie de criterios que pueden utilizarse para priorizar la publicación de algunos conjuntos de datos antes que otros.



Calidad del dato



Estrategia de publicación



Negocio

Reto a solucionar

Desarrollar los criterios para "conjuntos de datos de alto valor" teniendo en cuenta la probable reutilización de datos abiertos y ayudar a los gobiernos a entender qué conjuntos de datos priorizar para su publicación.

Implementación

Las características de los "conjuntos de datos de alto valor" se pueden observar desde tres perspectivas: reutilización, valor para propietarios de datos, valor para reutilizadores.

Reutilización:

- Los datos de alto valor deben alcanzar por lo menos 3 estrellas en el esquema de Tim Berners-Lee.

Valor para el propietario de los datos

Un conjunto de datos puede considerarse de alto valor cuando se cumplen uno o más de los siguientes criterios:

- Compartirlo contribuye a la transparencia.
- La publicación está sujeta a una obligación legal.
- Los datos se relacionan directa o indirectamente con su tarea pública.
- Compartirlo ayuda con la reducción de costos.

Valor para los reutilizadores

El valor de un conjunto de datos depende principalmente de su uso y su potencial para ser reutilizado, lo que puede conducir a la generación de actividad empresarial. El potencial del conjunto de datos se define por:

- El tamaño y la dinámica del público objetivo;
- El número de sistemas o servicios que podrían utilizar el conjunto de datos.

Los conjuntos de datos que contribuyen a la transparencia tienen un fuerte impacto social y el interés de los reutilizadores por estos conjuntos de datos es alto.

Involucrarse con los reutilizadores

Es importante involucrarse directamente con los reutilizadores para entender el valor de los conjuntos de datos. Recomendaciones:

- Establecer un canal de comunicación
- Utilizar herramientas colaborativas.

Ejemplo de implementación

El Grupo de datos abiertos de la Red de Entidades Locales por la Transparencia y la Participación Ciudadana de la FEMP establece una serie de conjuntos de datos mínimos a publicar por parte de Entidades Locales:

- Agenda (Municipal y Alcaldía) - Sector público
- Agenda de Actividades - Cultura y ocio
- Aparcamiento público - Transporte
- Avisos, Sugerencias, Quejas y Reclamaciones - Sociedad y bienestar
- Bicicleta pública (estaciones, tiempo real, itinerarios, usuarios, usos) - Transporte
- Calidad del aire - Medio ambiente
- Callejero (Viales, Numeraciones, Tramero, etc.) - Urbanismo e infraestructuras
- Censo de locales y actividades. Y Licencia de apertura de locales - Comercio
- Contaminación acústica (día, tarde, noche, industrial, tráfico rodado) - Medio ambiente
- Contrataciones, licitaciones y proveedores de servicios municipales - Sector público
- Convenios y contratos asociados - Sector público
- Deuda del Ayuntamiento - Economía
- Equipamientos municipales - Sociedad y bienestar. Cultura y Ocio
- Estado Tráfico en tiempo real (cortes, obras, intensidad, predicciones, ...) e histórico - Transporte
- Instalaciones deportivas de acceso público (oferta y asistencia). Y unidades deportivas (pistas) y servicios - Deporte
- Lugar de interés turístico (edificios de interés, edificios históricos, monumentos, etc.) - Cultura y ocio
- Padrón (Actual, Histórica, ...) - Demografía
- Presupuesto anual (son varios datasets) - Hacienda
- Puntos wifi público - Infraestructuras
- Transporte urbano - Transporte

En el anexo de "[Datos abiertos . Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar](#)" publicado por Grupo Datos Abiertos de la FEMP podemos encontrar un análisis mínimo de cada conjunto de datos, existiendo un fichero de detalle, con más información, y con el análisis de las diversas ciudades y lo que contienen al respecto.

Impacto



Reutilización de los datos



Confianza entre las partes



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [Tendencias y buenas prácticas en la implementación de políticas de datos abiertos](#)
- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos. Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar.](#)
- 2012 datos.gob.es [ES]: [Guía de aplicación del Real Decreto 1495/2011](#)
- 2013 Nicolas Loozen [EN]: [Good practices for identifying high value datasets and engaging with reusers: the case of public tendering data.](#)
- 2015 Makx Dekkers [EN]: [How good is good enough?](#)
- 2013 Anneke Zuidervijk, Iryna Sussha, Yannis Charalabidis, Peter Parycek, Marijn Janssen [EN]: [Specific Critical Success Factors for Open Data Publication and Use](#)
- Open Data Institute white paper [EN]: [How to prioritise open data to drive global development](#)
- Open Data Institute guidance [EN]: [Engaging with reusers](#)



Publicador

Proporcionar los datos en múltiples formatos (1/2)



Proporcionar datos en más de un formato reduce los costos incurridos en la conversión del formato de los datos. También minimiza la posibilidad de introducir errores en el proceso de transformación. Si muchos usuarios necesitan transformar los datos a un formato específico, publicar los datos en ese formato desde el principio ahorra tiempo y dinero y evita errores en la conversión. Por último, se aumenta el número de herramientas y aplicaciones que pueden procesar los datos.



Estrategia de publicación



Formatos

Reto a solucionar

Hacer que tantos usuarios como sea posible sean capaces de utilizar los datos sin tener que transformarlos primero al formato en el que se necesitan.

Implementación

Se debe considerar los formatos de datos más probables y las alternativas que probablemente serán más útiles en el futuro. Los publicadores deben equilibrar el esfuerzo requerido para hacer que los datos estén disponibles en muchos formatos con el costo de hacerlo, pero proporcionar al menos una alternativa aumentará considerablemente la usabilidad de los datos. Para hacer que los datos estén disponibles en más de un formato se puede utilizar la negociación de contenido como se describe en la buena práctica correspondiente.

Ejemplo de implementación

Se dispone de una distribución JSON del conjunto de datos de paradas de autobús. En el siguiente ejemplo, la propiedad dcat: distribution se utiliza para asociar el dataset stops-2015-05-05 con sus dos distribuciones: stops-2015-05-05.csv y stops-2015-05-05.json.

```

:stops-2015-05-05
  a dcat:Dataset ;
  dcat:distribution :stops-2015-05-05.csv ;
  dcat:distribution :stops-2015-05-05.json
.

:stops-2015-05-05.csv
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv> ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/>
.

:stops-2015-05-05.json
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.json> ;
  dct:title "JSON distribution of stops-2015-05-05 dataset" ;
  dct:description "JSON distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "application/json" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/>
.

```

Validación

Comprobar si el conjunto de datos completo está disponible en más de un formato.



Publicador

Proporcionar los datos en múltiples formatos (2/2)



Impacto



Reutilización de los datos



Procesamiento de los datos



Estrategia de publicación

Referencias

- European data portal [EN]: Data formats: [Most common data formats used](#)
- 2015 W3C [EN]: [R-FormatMultiple](#)



Publicador

Utilizar la negociación de contenido para publicar datos disponibles en varios formatos



Es posible servir datos en una página HTML mezclando datos legibles por el usuario y procesables por máquinas, usando RDFs, por ejemplo. Sin embargo, como demuestra la Arquitectura de la Web y el DCAT, un recurso, como un conjunto de datos, puede tener muchas representaciones. Los mismos datos podrían estar disponibles en diferentes formatos como son JSON, XML, RDF, CSV y HTML. Estas representaciones múltiples pueden estar disponibles a través de una API, pero deben estar disponibles desde la misma URL usando la negociación de contenido para devolver la representación apropiada (lo que DCAT llama una distribución). Los URI específicos pueden usarse para identificar representaciones individuales de los datos directamente, pasando por alto la negociación de contenido.



Identificación



Estrategia de publicación



Documentación/
Catalogación

Reto a solucionar

La negociación de contenido permitirá que diferentes recursos o diferentes representaciones del mismo recurso se sirvan de acuerdo al tipo de solicitud hecha por el cliente.

Implementación

Un posible enfoque para la implementación es configurar el servidor Web para tratar la negociación de contenido del recurso solicitado.
El formato específico de la representación del recurso solicitado puede ser accedido por el URI o por la cabecera Content-type de la petición HTTP.

Ejemplo de implementación

Las diferentes representaciones del dataset de paradas de bus pueden ser servidas de acuerdo con el tipo de contenido especificado de la solicitud HTTP:
Si se usase cURL para obtener el contenido de <http://data.mycity.example.com/transport/dataset/bus/stops> que se encuentra en formato CSV y en formato JSON-LD, las peticiones tendrían la siguiente forma:

```
curl -L -H "Accept: text/csv" http://data.mycity.example.com/transport/dataset/bus/stops
```

```
curl -L -H "Accept: application/ld+json" http://data.mycity.example.com/transport/dataset/bus/stops
```

Validación

Comprobar que existen diferentes representaciones disponibles del mismo recurso e intentar obtenerlas especificando el contenido deseado en el encabezado de la solicitud HTTP.

Impacto



Reutilización de los datos



Legibilidad de los datos



Acceso a los datos

Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2008 W3C [EN]: [Best Practice Recipes for Publishing RDF Vocabularies](#)
- W3C [EN]: [R-FormatMachineRead](#)
- W3C [EN]: [R-FormatMultiple](#)



Publicador

Publicar datos estadísticos en formato Linked Data (1/2)



La publicación de datos estadísticos como datos enlazados siguiendo las bases RDF Data Cube vocabulary del W3C supone la expresión de los datos de una manera estandarizada, legible por máquinas, así como la identificación de los metadatos recomendados para describir los conjuntos de datos.



Formatos



Calidad del dato



Identificación



Estrategia de publicación

Reto a solucionar

Los datos estadísticos se publican actualmente en una gran variedad de formatos y estándares, lo que no permite el enlace entre distintos conjuntos de datos. Los datos estadísticos se utilizan normalmente como base para la predicción, planificación y ajustes de políticas, y por lo tanto tienen un impacto significativo en la sociedad (para los ciudadanos, las empresas y los gobiernos). El proceso de la recolección y el seguimiento de indicadores socioeconómicos puede mejorarse considerablemente si los datos producidos por organizaciones gubernamentales tales como oficinas de estadística, bancos nacionales, servicios de empleo, etc., se publican en formato de datos enlazados (Linked Data).

Implementación

Linked Data ha abierto nuevas posibilidades y perspectivas para que las organizaciones gubernamentales abran datos e intercambien información. Los datos se consideran abiertos si están abiertos técnicamente (disponibles en un formato estándar legible por máquina, lo que significa que pueden ser recuperados y procesados de manera significativa por una aplicación informática) y abiertos legalmente (con licencia explícita que permita el uso comercial y no comercial y Reutilización sin restricciones).

El enfoque Linked Data permite enlazar los conjuntos de datos mediante referencias a conceptos comunes. Un conjunto de datos se representa en forma de grafo, utilizando RDF como lenguaje de propósito general. En el proceso de publicación de datos enlazados se intervienen un conjunto de actividades como la extracción, transformación, validación, exploración y publicación de conjuntos de datos en formato RDF procedentes de diferentes fuentes (por ejemplo, bases de datos). Los conjuntos de datos en RDF listos para usarse pueden almacenarse localmente o ser registrados en un catálogo de metadatos.

En 2014 fue publicado el RDF Data Cube Vocabulary por el W3C como una recomendación para publicar datos multidimensionales en la Web.

Esta buena práctica se basa en un conjunto de herramientas para automatizar el proceso de extracción y publicación de datos. La comunidad de investigación de la UE puso a disposición muchas herramientas de código abierto para publicar los datos estadísticos en formato de datos enlazados, por ejemplo:

- LOD2 [Statistical Workbench](#):
- [Open Cube](#)

Ejemplo de implementación

- Italia [LOD ISTAT](#)
- Italia [LinkedStat](#)
- Reino Unido [Scottish Government Statistics](#)
- Finlandia [Semangtic hri.fi](#)
- Republica Checa [Publikace dat statistických ročenek ve standardu otevřených dat](#)

Impacto



Reutilización de los datos



Interoperabilidad



Procesamiento de los datos



Publicador

Publicar datos estadísticos en formato Linked Data (2/2)



Referencias

- 2016 Esquema Nacional de Interoperabilidad [ES]: [Guía de aplicación de la Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información](#)
- 2014 European data portal [EN]: Training Module [Introduction to Linked Data](#)
- 2014 European data portal [EN]: Training Module [Introduction to RDF & SPARQL](#)
- 2014 Valentina Janev, Institut Mihajlo Pupin [EN]: [Publishing and Consuming Linked Open Data with the LOD Statistical Workbench](#).
- 2014 Agenzia per l'Italia Digitale (Giorgia Lodi, Antonio Maccioni), Istituto Nazionale di Statistica (Monica Scannapieco, Mauro Scanu, Laura Tosco) [EN]: [Publishing Official Classifications in Linked Open Data](#).



Reutilización de vocabularios, preferiblemente estándares (1/2)



La descripción semántica de recursos a través del uso de vocabularios ya utilizados facilita el consenso en las comunidades, aumenta la interoperabilidad y reduce las redundancias, fomentando así la reutilización de los propios datos. En particular, el uso de vocabularios comunes para metadatos (especialmente los metadatos estructurales, de procedencia, de calidad y de versiones) ayuda a la comparación y procesamiento automático de datos y metadatos. Además, referirse a los códigos y términos de las normas ayuda a evitar la ambigüedad y los conflictos entre elementos con valores similares.



Calidad del dato



Documentación/
Catalogación

Reto a solucionar

Mejorar la interoperabilidad y el consenso entre los productores de datos y los consumidores.

Implementación

La sección "Vocabularios" de las prácticas recomendadas para publicar datos enlazados del W3C proporciona una orientación sobre el descubrimiento, la evaluación y la selección de vocabularios existentes.

Organizaciones como el OGC, ISO, W3C, OMM, bibliotecas y servicios de datos de investigación, etc. proporcionan listas de códigos, terminologías y vocabularios de Linked Data que pueden ser usados por todos. Un punto clave es asegurarse de que el conjunto de datos, o su documentación, proporcione suficiente contexto para que los consumidores de datos puedan recuperar y explotar el significado estándar de los valores. En el contexto de la Web, el uso de URIs para recursos de vocabulario estándares es una forma eficiente de hacerlo.

Ejemplo de implementación

La Biblioteca del Congreso publica listas de lenguajes ISO 639 como datos:

```
:stops  
dct:language <http://id.loc.gov/vocabulary/iso639-1/en> .
```

Google mantiene la Especificación de canal Tránsito General que define un formato para publicar datos de transporte público. Este formato se basa en un conjunto de campos como `route_short_name` o `route_type` que están cuidadosamente definidos y expuestos a la constante retroalimentación de la comunidad para facilitar el consenso. Las definiciones incluyen especificaciones de valores codificados, como las utilizadas con `route_type`:
0 - tranvía, tranvía, tren ligero. Cualquier sistema de tren ligero o de nivel de calle dentro de un área metropolitana.
1 - Metro, Metro. Cualquier sistema de metro en un área metropolitana.
2 - Carril. Se utiliza para viajes interurbanos o de larga distancia.

El dataset [Callejero de Zaragoza](#), publicado por el portal de datos abiertos de Zaragoza, utiliza [este vocabulario](#) para la representación del callejero.

Validación

El uso de repositorios de vocabulario como el repositorio Linked Open Vocabularies o las listas de servicios mencionados en las Buenas Prácticas específicas de esta tecnología, como las Buenas Prácticas para publicar datos vinculados [LD-BP] o el Contexto inicial básico de RDFa y JSON-LD, permiten comprobar que las clases, propiedades, términos, elementos o atributos utilizados para representar un conjunto de datos no replican los definidos por los vocabularios utilizados para otros conjuntos de datos.

Se deben comprobar si los términos o códigos del vocabulario que va ser utilizado se definen en una organización de desarrollo de estándares como IETF, OGC & W3C etc., o son publicados por una autoridad adecuada como una agencia gubernamental.

Impacto



Reutilización de los datos



Interoperabilidad



Procesamiento de los datos

Referencias

- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos. Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar.](#)
- European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2014 European data portal [EN]: Training Module [Designing and Developing vocabularies in RDF](#)
- 2015 W3C [EN]: [R-MetadadataStandardized](#)
- 2015 W3C [EN]: [R-MetadadataDocum](#)
- 2015 W3C [EN]: [R-QualityComparable](#)
- 2015 W3C [EN]: [R-VocabOpen](#)



Publicador

Uso de formatos de datos estándares y legibles por máquinas (1/2)



A medida que los datos se hacen más omnipresentes y los conjuntos de datos se hacen más grandes y complejos, el procesamiento automatizado se hace cada vez más crucial. El envío de datos en un formato que no sea legible por máquinas impone severas limitaciones a la utilidad de los datos. Los datos se vuelven útiles cuando se han procesado y transformado en información. Hay que tener en cuenta que hay una distinción importante entre los formatos que pueden ser leídos y editados por los seres humanos utilizando un equipo informático y los formatos que son legibles por máquinas. Este último término implica que los datos son fácilmente extraídos, transformados y procesados por un ordenador.

El uso de formatos de datos no estándar es costoso e ineficiente, y los datos pueden perder sentido a medida que se transforman. Por el contrario, los formatos de datos estandarizados permiten la interoperabilidad, así como usos futuros, tales su mezcla o visualización, muchos de los cuales no se pueden anticipar cuando los datos se publican por primera vez.



Formatos

Reto a solucionar

Conseguir que las máquinas sean capaces de leer y procesar los datos publicados en la web más fácilmente y que los humanos puedan utilizar herramientas computacionales fácilmente accesibles para trabajar con los datos.

Implementación

Se pretende hacer que los datos estén disponibles en un formato estandarizado y legible por máquinas que sea fácilmente analizable. Por ejemplo: CSV, XML, HDF5, JSON y RDF serializado como RDF/XML, JSON-LD o Turtle.

Ejemplo de implementación

Se sabe que los datos tabulados se usan comúnmente en la web y se decide usar CSV como formato de datos para una de las distribuciones del dataset de paradas de bus. Para facilitar el procesamiento de datos, se utiliza el Model for Tabular Data and Metadata on the Web para publicar la distribución en formato CSV (stops-2015-05-05.csv). El ejemplo siguiente presenta un fragmento de la distribución CSV:

```
Identifier,Name,Description,Latitude,Longitude,ZONE,URL
345,Castle Avenue,Sunset Drive,-3.731862,-
38.526670,x20,http://data.mycity.example.com/transport/road/bus/stop/id/345
483,Main Street,Lily Park,-3.731541,-
38.535157,x20,http://data.mycity.example.com/transport/road/bus/stop/id/483
```

El ayuntamiento de Alcobendas pone a disposición de los usuarios el dataset [Horarios de cercanías](#) que tiene distribuciones disponibles en los formatos CSV, JSON y XLSX. Un fragmento de la distribución en formato JSON es:

```
{
  "trip_id": "5_I13-5_I13_11:32:00_2_46_5__4_A__",
  "arrival_time": "12:24:00",
  "stop_sequence": "13",
  "stop_id": "par_5_36",
  "departure_time": "12:24:00"
}
```

Validación

Comprobar si alguno de los formatos publicados es legible por máquinas.



Publicador

Uso de formatos de datos estándares y legibles por máquinas (2/2)



Impacto



Reutilización de los datos



Procesamiento de los datos

Referencias

- 2017 datos.gob.es [ES]: [Buenas prácticas en la apertura de datos a lo largo del mundo](#)
- European data portal [EN]: Data formats: [Most common data formats used](#)
- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos . Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar.](#)
- 2017 European Data Portal [EN]: [Practical Guide Preparing data](#)
- 2015 W3C [EN]: [R-FormatMachineRead](#)
- 2015 W3C [EN]: [R-FormatStandardized](#)
- 2015 W3C [EN]: [R-FormatOpen](#)



Publicador Reutilizador

Estándares para datos geospaciales (1/2)



Los organismos de la administración pública deben trabajar conjuntamente con empresas de arquitectura, ingeniería y construcción, así como propietarios de edificios, corredores de bolsa, proveedores de componentes, operadores, aseguradores, inspectores, arrendatarios, compañías financieras, bomberos, servicios sociales o servicios de salud. Para gran parte de la información del sector público (PSI en inglés), la localización es crítica. Por lo tanto, es esencial que los datos geográficos y geospaciales se compartan usando estándares, de manera que sea más factible su reutilización. La mayoría de los estándares relevantes para los datos geospaciales son desarrollados por el Open Geospatial Consortium (OGC).



Formatos

Reto a solucionar

El reto es asegurar que información de diferentes fuentes relacionada con la misma ubicación se pueda utilizar de manera conjunta. Esto es particularmente difícil porque la localización se puede definir de muchas maneras diferentes: nombre (en diferentes idiomas y o diferentes abreviaturas), coordenadas, límites, nombres de los distritos administrativos, código NUTS, planos de planta, puntos, centroides, polígonos y rasters, mapas de metro, paradas de autobús, series temporales, direcciones izquierda y derecha, etc. El objetivo es hacer que todos estos datos estén disponibles como datos abiertos, siguiendo estándares abiertos y modelos de datos abiertos. La precisión, incertidumbre, procedencia, derechos y control de acceso a la información son factores críticos. Dada toda esta complejidad, el desarrollo de un software que implique datos de ubicación puede ser difícil. El requisito de que los datos de localización y los servicios de localización sean compatibles, reutilizables e interoperables hace que la tarea sea aún más difícil.

Otro desafío importante es que las tecnologías Web están evolucionando por lo que los estándares OGC y las arquitecturas de información de datos geospaciales también necesitan evolucionar.

Implementación

La información geográfica y de geolocalización del sector público puede proporcionar muchos conjuntos de datos de manera eficiente e interoperable mediante el uso de estándares, especialmente mediante el uso de normas desarrolladas por el OGC. Las normas OGC e ISO (como WFS, WMS, GML, IndoorGML, CityGML y SOS) garantizan un acceso estandarizado a toda la información del sector público con características espaciales.

Ejemplo de implementación

Las normas de OGC son ampliamente implementadas en todo el mundo y constituyen la base de muchas industrias críticas y actividades gubernamentales. Véase, por ejemplo:

[Inforshare de la región de Helsingland](#)

Y la aplicación checa de la Directiva [INSPIRE](#)

La infraestructura de Datos Espaciales ofrecida por el Ayuntamiento de Gijón ([IDE Gijón](#)) se basa en los diversos servicios OGC. [En este enlace](#) podemos encontrar información y accesos a los servicios WMS, WFS, WCS y CSW proporcionados por el IDE Gijón.

Validación

Comprobar el uso correcto de tecnologías OGC (<http://www.opengeospatial.org/docs/is>) y estándares W3C (<https://www.w3.org/standards/>)

Impacto



Reutilización de los datos



Interoperabilidad



Procesamiento de los datos



Referencias

- 2013 Universidad Técnica de Ingeniería Civil de Bucarest [EN]: [Free our maps](#)
- 2013 Share-PSI 2.0 [EN]: [Location Track](#)
- 2016 W3C [EN]: [Spatial Data on the Web Best Practices](#)



Publicador

Proporcionar subconjuntos para conjuntos de datos grandes



Los conjuntos de datos grandes pueden ser difíciles de mover de un lugar a otro. También puede ser un inconveniente para los usuarios el almacenarlos o analizarlos. Los usuarios no deberían tener que descargar un conjunto de datos completo si solo necesitan un subconjunto de él. Por otra parte, las aplicaciones Web que aprovechan conjuntos de datos grandes tendrán un mejor rendimiento si sus desarrolladores pueden aprovechar la "carga perezosa", trabajando con piezas más pequeñas de un todo y tirando de nuevas piezas sólo cuando sea necesario. La capacidad de trabajar con subconjuntos de datos también permite que el procesamiento sin conexión funcione de manera más eficiente. En particular se benefician las aplicaciones en tiempo real, ya que pueden actualizarse más rápidamente.



Calidad del dato



Estrategia de publicación

Reto a solucionar

Hacer que los seres humanos y las aplicaciones puedan acceder a subconjuntos de un conjunto de datos mayor. Se pretende que los conjuntos de datos estáticos que los usuarios consideran demasiado grandes se puedan descargar en fragmentos más pequeños. Las API crearán segmentos o subconjuntos filtrados de los datos disponibles, la granularidad dependerá de las necesidades del dominio y las demandas de rendimiento en una aplicación web.

Implementación

Hay que considerar los casos de un uso separado de un conjunto de datos y determinar qué tipos de subconjuntos pueden ser más útiles. Una API es generalmente el enfoque más flexible para servir a subconjuntos de datos, ya que permite la personalización de los datos que se transfieren, lo que hace que los subconjuntos disponibles sean mucho más propensos a proporcionar los datos necesarios - y pocos datos innecesarios - para cualquier situación dada. La granularidad debe ser adecuada para las velocidades de acceso a aplicaciones Web.

Otra forma de trocear un conjunto de datos es simplemente dividirlo en unidades más pequeñas y hacer que esas unidades individualmente disponibles para su descarga o visualización.

También puede ser útil marcar un conjunto de datos para que las secciones individuales a través de los datos (o incluso piezas más pequeñas, si los casos de uso esperado lo justifican) se pueden procesar por separado. Una forma de hacerlo es definir "slices" (o lonchas) con el vocabulario RDF Data Cube.

Ejemplo de implementación

La agencia de transporte de MyCity ha estado recopilando datos detallados sobre el uso de los pasajeros por varios años. Se trata de un conjunto de datos muy amplio, que contiene valores para el número de pasajeros por tipo de tránsito, ruta, vehículo, conductor, parada de entrada, salida, tiempo de entrada, etc. Permitiendo descargar varios subconjuntos de los datos. La gente que dirige cada sistema de tránsito desea sólo los datos para su modo de tránsito, los planificadores de la ciudad sólo quieren el número de entradas y salidas en cada parada, la oficina de presupuesto de la ciudad sólo quiere los números para los diversos tipos de tickets vendidos, y otros organismo quieren diferentes subconjuntos. La agencia creó un sitio Web donde los usuarios pueden seleccionar qué variables son de interés para ellos, establecer rangos en algunas variables y descargar sólo el subconjunto que necesitan.

Validación

Comprobar que el conjunto de datos completo se puede recuperar haciendo varias solicitudes que recuperan unidades más pequeñas.

Impacto



Reutilización de los datos



Procesamiento de los datos



Acceso a los datos

Referencias

- 2015 W3C [EN]: [R-Citable](#)
- 2015 W3C [EN]: [R-GranularityLevels](#)
- 2015 W3C [EN]: [R-UniqueIdentifier](#)
- 2015 W3C [EN]: [R-AccessRealTime](#)



GOBIERNO DE ESPAÑA

MINISTERIO DE HACIENDA Y FUNCIÓN PÚBLICA

MINISTERIO DE ENERGÍA, TURISMO Y AGENDA DIGITAL

red.es

Iniciativa **aporta**



Publicador

Uso de datos de transporte público (1/2)



Una de las principales prioridades de la UE es promover un transporte inteligente, ecológico e integrado para el beneficio de todos los ciudadanos, la economía y la sociedad. Uno de los resultados esperados es la elaboración de nuevos modelos de negocio para el transporte público a través de tecnologías (tales como servicios de TI y de aplicaciones) e innovaciones sociales, teniendo en cuenta las posibles barreras sociales y demográficas.

En este sentido, la información de transporte público (horarios, interrupciones de servicios, paradas, accesibilidad, etc) se considera como un dato de alto valor, pieza esencial para posibilitar esta innovación, garantizando servicios más eficientes, fiables y atractivos tanto para los operadores como para los clientes. Por lo general, esta información se muestra en las paradas (en tabloneros de anuncios o pantallas electrónicas) por lo que la información se considera no sensible y pública, y el problema de obtener esta información a través de la Web debe ser sólo un problema técnico. Por lo tanto, el alto valor de este conjunto de datos para toda la sociedad, en contraste con el esfuerzo mínimo que supone publicar la información ya existente hace que los datos de transporte público sean una prioridad absoluta para las iniciativas de reutilización de datos abiertos y PSI.

Aunque los servicios de transporte pueden ser dirigidos por empresas privadas, los gobiernos deben garantizar la publicación de esta información. Esto puede hacerse a través de acciones legales como la emisión de políticas específicas o la adición de cláusulas en la contratación pública.

Todas las partes interesadas se beneficiarán: una mejor experiencia de usuario para los usuarios; ciudades más verdes gracias al uso de transporte público (o colectivo); y más eficiencia sin costes adicionales importantes.



Estrategia de publicación



Negocio



Comunidad

Reto a solucionar

Aunque la mayor parte de esta información no es sensible y se considera de alto potencial para su reutilización, no siempre está expuesta públicamente. La mayoría de las empresas de transporte ya gestionan la información operativa (horarios, estado del servicio, etc.) por medios electrónicos. Por lo tanto, el costo de liberar la información abiertamente no debe ser demasiado alto. Aunque el beneficio potencial es para toda la sociedad (incluido el gobierno y la empresa de transporte), muchas de estas empresas son reacias a publicar los datos, incluso teniendo solicitudes directas de los gobiernos locales, regionales y nacionales.

Implementación

Los gobiernos locales deben hacer un esfuerzo para publicar la información del transporte público en formatos fáciles de usar y de lectura mecánica, dirigidos principalmente a empresas y desarrolladores que podrán crear nuevos servicios o productos. Si el gobierno no tiene acceso total ni control sobre la información, deberá convencer a la compañía de transporte para que la haga pública.

Esto beneficiaría a muchas partes interesadas:

- Los municipios y gobiernos que sigan las estrategias de "Smart City" están interesados en reducir el tráfico en la ciudad, mejorar el transporte público y fomentar su uso
- Los usuarios actuales del transporte público disfrutarían de un mejor servicio
- Las empresas privadas podrían producir nuevos productos y servicios para los viajeros

Ejemplo de implementación

Algunas iniciativas de datos abiertos donde se implementa esta BP son:

- [Barcelona](#)
- [Bilbao](#)
- [Gijón](#)
- [Cáceres](#)
- [Madrid](#)
- [Gipuzkoa](#)
- [Granollers](#)
- [La Palma](#)
- [Las Palmas de Gran Canaria](#)
- [Málaga](#)
- [Sabadell](#)
- [Santander](#)
- [Terrasa](#)
- [Tenerife](#)
- [Valencia](#)
- [Zaragoza](#)



Publicador

Uso de datos de transporte público (2/2)



Impacto



Estrategia de publicación

Referencias

- 2016 European data portal: Analytical Report [Open Data and Cities](#)
- 2014 Samos Workshop Talk [EN]: [Open Traffic Information Standard & Experimentation for Enhanced Services](#) (PDF)
- 2014 Samos Workshop Talk [EN]: [Public Transport Data in the City of Gijón](#) (PDF)
- 2013 Krems Workshop Session [EN]: [OpenMove: How Trentino opened public transportation data and benefitted of a mobile ticketing solution for free](#) (PDF)



Utilizar representaciones de datos deslocalizadas (1/2)



Los valores de datos que son legibles por máquinas y que no están especificados en ningún idioma en particular son más duraderos y menos predispuestos a interpretaciones erróneas. Aspectos como fechas, monedas y números pueden ser representados de forma similar, pero tienen diferentes significados en diferentes lugares. Por ejemplo, la fecha 4/7 se puede leer como el 7 de abril o el 4 de julio, dependiendo de dónde se crearon los datos. Mediante el uso de un formato neutral, los sistemas evitan la necesidad de establecer reglas específicas de intercambio que varían según el idioma o la ubicación del usuario.



Formatos



Calidad del dato



Documentación/
Catalogación

Reto a solucionar

Hacer que las personas y los agentes software sean capaces de interpretar el significado de textos que representan fechas, horas, monedas y números, etc.

Implementación

Los formatos de serialización de datos más comunes son regionalmente neutrales. Por ejemplo, los tipos de esquema XML, como `xsd:integer` y `xsd:date`, están diseñados para el intercambio de datos que no esté limitado a una región. El uso de ese tipo de representaciones permite que los valores de los datos se procesen y extraigan con precisión y no de lugar a una mala interpretación. También permite que los datos se presenten en el formato más cómodo para el consumidor de cualquier región. Por ejemplo, en lugar de almacenar "€ 2000,00" como una cadena de texto, se prefiere intercambiar una estructura de datos como:

```
"price" {
  "value": 2000.00,
  "currency": "EUR"
}
```

Algunos conjuntos de datos contienen valores que no están o no se pueden representar en un formato de este tipo de formatos. Esto se particulariza para cierto en los valores de texto de un lenguaje específico. Para cada campo de este tipo de datos, que puede contener texto en cierto idioma o lenguaje, debería haber una etiqueta de idioma asociada para indicar el idioma y la configuración regional de los datos. Esta información de configuración regional se puede utilizar para analizar los datos o para garantizar la presentación y el procesamiento adecuados del valor por parte del consumidor.

Ejemplo de implementación

Legibilidad por humanos:

La siguiente [página de ejemplo](#) contiene una descripción legible por humanos de un conjunto de datos

Legibilidad por máquinas:

El siguiente ejemplo muestra los metadatos legibles por máquina para el dataset de paradas de bus con la inclusión de metadatos de parámetros de configuración regional. La propiedad `dct:language` se utiliza para declarar los idiomas en los que se publica el conjunto de datos. Si el conjunto de datos está disponible en varios idiomas, se utilizan varios valores para esta propiedad. La propiedad `dct:conformsTo` se utiliza para especificar la norma adoptada para los formatos de fecha y hora:

```
:stops-2015-05-05
a dcat:Dataset ;
dct:title "Bus stops of MyCity" ;
dcat:keyword "transport","mobility","bus" ;
dct:issued "2015-05-05"^^xsd:date ;
dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
dct:spatial <http://sws.geonames.org/3399415> ;
dct:publisher :transport-agency-mycity ;
dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
dcat:theme :mobility ;
dcat:distribution :stops-2015-05-05.csv ;
dct:language <http://id.loc.gov/vocabulary/iso639-1/en> ,
<http://id.loc.gov/vocabulary/iso639-1/pt> ;
dct:conformsTo <http://www.iso.org/iso/home/standards/iso8601.htm>
```

El portal de transparencia y datos abiertos del Ayuntamiento de Valencia permite a los usuarios [descargar el Catálogo en formato RDF-DCAT](#). Se puede observar que se declaran los idiomas en que van a usar los conjuntos de datos bajo la propiedad `dc:language`:

```
...
<dc:language>es</dc:language>
<dc:language>ca</dc:language>
<dc:language>en</dc:language>
...
```



Publicador

Utilizar representaciones de datos deslocalizadas (2/2)



Validación

Comprobar que los valores de datos sensibles a la configuración regional están representados en un formato deslocalizado y si esto no es posible, que se proporcionen metadatos al respecto.

Impacto



Legibilidad de los datos



Interoperabilidad



Procesamiento de los datos

Referencias

- 2015 W3C [EN]: [R-FormatLocalize](#)
- 2015 W3C [EN]: [R-MetadataAvailable](#)
- 2015 W3C [EN]: [R-GeographicalContext](#)
- 2015 W3C [EN]: [R-FormatMachineRead](#)



Reutilizador

Seguir los términos de licencia



La concesión de licencias proporciona un marco legal para usar el trabajo de otra persona. Adherirse a los requisitos del editor original permite mantener una buena relación entre el editor y el usuario de esos datos. No es necesario preocuparse por acciones legales siempre que se sigan los requisitos establecidos por el editor original. Entender la licencia inicial ayudará a determinar qué licencia se selecciona tras la reutilización.



Licencias

Reto a solucionar

Se pretende conseguir que los editores de datos puedan confiar en que su trabajo se está reutilizando de acuerdo con sus requisitos de licencia, lo que hará que estén mas dispuestos a continuar publicando datos. Los reutilizadores serán capaces de licenciar adecuadamente sus trabajos derivados.

Implementación

Leer la licencia original y cumplir con sus requisitos. Si la licencia especifica otras para todo aquel trabajo derivado, se deberá escoger una para cumplir con este requisito. Si no se indica ninguna licencia, se debe poner en contacto con el editor original.

Ejemplo de implementación

Si un conjunto de datos que está utilizando está licenciado bajo la licencia [Creative Commons Attribution 3.0](#), tendrá que cumplir con los términos especificados para ésta.

En el apartado [Cómo utilizar los datos – Condiciones de uso](#) de la página web del proyecto de datos abiertos Generalitat de Catalunya se especifica las condiciones a seguir para reutilización de la información publicada por ellos.

Validación

Leer la licencia original y comprobar que el uso de los datos no viola ninguno de los términos.

Impacto



Confianza entre las partes



Reutilización de los datos

Referencias

- European data portal [EN]: Short overview of licences: [Introduction to licences for Open Data and explanation to the most often used licences](#)
- 2014 European data portal [EN]: Training Module [Data & metadata licensing](#)
- 2014 datos.gob.es [ES]: [Decálogo del Reutilizador de Datos Abiertos del Sector Público](#)
- 2015 W3C [EN]: [R-LicenseAvailable](#)
- 2015 W3C [EN]: [R-LicenseLiability](#)



Publicador

Proporcionar datos actualizados (1/2)



La disponibilidad de datos en la Web debería coincidir con la fecha de creación de datos o la fecha en la que fueron recogidos. Una sincronización cuidadosa entre la frecuencia de actualización y la publicación de datos fomenta la confianza de los consumidores y la reutilización de los datos.



Calidad del dato



Estrategia de publicación

Reto a solucionar

Disponer de las versiones más recientes de los datos en la Web, a la vez que se hacen públicos por cualquier otro canal. Cuando se disponga de nuevos datos se publicarán en la Web lo antes posible.

Implementación

Las nuevas versiones de los conjuntos de datos se pueden publicar en la Web de manera regular y programada. Publicar en la Web debe ser una parte del proceso de lanzamiento de nuevas versiones de los datos. Hacer de la publicación en la Web un elemento entregable en el proceso y asignar a un individuo como responsable de la tarea puede ayudar a evitar que los datos se queden obsoletos. Para cumplir con las expectativas de actualización de los consumidores es recomendable publicar la frecuencia de publicación esperada como metadatos, que sean legibles, tanto por humanos como por máquinas.

Ejemplo de implementación

Supongamos que la frecuencia de actualización del conjunto de datos de paradas de autobuses es anual. Para describir la frecuencia con la que se agregan nuevos datos al conjunto de datos, se puede utilizar la propiedad `dct:accrualPeriodicity`. Se crea una nueva versión del conjunto de datos (stops-2016-05-05) para reflejar el programa de actualización de los datos. Es importante tener en cuenta que las nuevas versiones pueden crearse antes de lo previsto, pero el editor debe asegurarse de que las versiones adicionales se publican en la Web tan rápidamente como las programadas.

```
:stops-2016-05-05
a dcat:Dataset ;
dct:title "Bus stops of MyCity" ;
dcat:keyword "transport","mobility","bus" ;
dct:issued "2016-05-05"^^xsd:date ;
...
dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
...
dct:isVersionOf :stops-2015-05-05 ;
pav:previousVersion stops-2015-12-17 ;
rdfs:comment "The bus stops dataset was updated to reflect the creation of new bus stops since the last update and to follow the update frequency" ;
owl:versionInfo "1.2" ;
pav:version "1.2"
```

En el dataset [Contaminación acústica: datos en tiempo real](#) del Portal de datos abiertos del Ayuntamiento de Madrid vemos, en la sección "Frecuencia de actualización", que posee una frecuencia de actualización diaria, por lo que la información que podemos encontrar en la distribución que se nos presenta será constantemente actualizada.

Validación

Comprobar que se ha indicado la frecuencia de actualización y que la copia publicada más recientemente en la Web no es anterior a la fecha prevista por la frecuencia de actualización indicada.



Proporcionar datos actualizados (2/2)



Impacto



Reutilización de los datos



Confianza entre las partes



Acceso a los datos

Referencias

- 2014 European data portal [EN]: Training Module [Introduction to Open Data Quality](#)
- 2015 W3C [EN]: [R-AccessUptodate](#)



Publicador

Uso de herramientas federadoras



Diferentes países han desarrollado herramientas federadoras para la recopilación de datos abiertos que permiten la publicación automática de los metadatos correspondientes a los conjuntos de datos publicados en los sitios web de cada entidad pública. Así se puede crear un índice global de información pública reutilizable que permite a los usuarios para localizar datos reutilizables sin necesidad de que estos conozcan el sitio web de la entidad pública que contiene los datos en los que están interesados.



Procedencia



Estrategia de publicación

Reto a solucionar

Hay un gran número de organismos públicos proveedores de datos y varios niveles de gobierno, por lo que sería muy difícil de implementar un sistema central que aglutine todos los datos abiertos del país y donde todo el mundo está autorizado a conectarse. Además, los propietarios de datos suelen ser bastante autónomos entre sí, por lo que no se puede implementar una solución técnica unificada en sus procesos internos. Otro aspecto a tener en cuenta es que los usuarios finales son muy variados por lo que es necesario un entendimiento común.

Implementación

Se implementará una solución federada, distribuida, escalable y capaz de integrar numerosos publicadores de datos. Los publicadores se integran a través de un formato de datos de salida común que plantea pocos requisitos sobre los procesos internos y utiliza un vocabulario común:

- Utilizar un portal de datos como un único punto donde federar todos los conjuntos de datos publicados de las diferentes entidades públicas. Acordar la organización de URIs persistentes, desde donde se puede acceder al origen de los datos. URIs asignadas deben identificar persistentemente lo mismo en el tiempo y el objeto identificado debe estar persistentemente disponible.
- Seleccionar una estructura o formato de datos intermedios. A medida que más y más sistemas se interconectan unos con otros, las soluciones estandarizadas son preferibles. Sin embargo, en algunos casos podría ser tan simple como un CSV con nombres de columnas acordados.
- Seleccionar un conjunto de herramientas para enviar los datos a un portal de publicación (por ejemplo, un recolector y un portal nacional, Portal Europeo de Datos Abiertos). Configurar las herramientas para recopilar los datos distribuidos que se publicarán en el portal.
- Designar una agencia responsable de mantener la estructura de datos. De ser necesario, debería establecerse una estructura de coordinación más sofisticada entre los diferentes niveles administrativos (estatales, regionales y locales).
- Establecer un marco jurídico y técnico que garantice que cada entidad pública federará sus conjuntos de datos en el portal nacional de datos de manera estándar.
- Monitorizar y apoyar el uso de las herramientas seleccionadas.

Ejemplo de implementación

Algunos ejemplos de la implementación de esta práctica son:

- España: [Catálogo Nacional de Datos Abiertos](#)
- Unión Europea: [European Data Portal](#)
- Estonia: [Catálogo de servicios públicos](#)
- Alemania: [Oficina de coordinación TI](#)

Impacto



Acceso a los datos



Estrategia de publicación

Referencias

- 2016 datos.gob.es [ES]: [Guía para organismos públicos.](#)
- 2015 Sebastian Sklarß, Lutz Rabe [EN]: [German XML for public administration "XÖV" tool chain in action](#)
- 2013 M^a Dolores Hernández Maroto [EN]: [A Federation Tool for Open Data Portals](#)
- 2013 Nikolaos Loutas, Makx Dekkers [EN]: [Implementing The DCAT Application Profile For Data Portals In Europe](#)
- 2013 Hannes Kivvet [EN]: [Estonian Metadata Reference Architecture](#)

ACCESO



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Publicador

Crear de un portal de Gobierno Abierto para compartir datos (1/2)



Los portales de datos contribuyen a la distribución de los datos abiertos porque proporcionan un punto de acceso y búsqueda para múltiples conjuntos de datos. Normalmente también actúan como un escaparate para la reutilización de datos y como punto de encuentro para la comunidad interesada.



Estrategia de publicación



Comunidad

Reto a solucionar

Las administraciones públicas están compuestas por múltiples organismos gubernamentales de tamaños diferentes y, a menudo, sus sistemas de información no son interoperables entre sí. Los conjuntos de datos se presentan en diferentes tamaños, a veces pertenecen sólo pertenecientes a pequeñas localidades, lo que los hace menos interesantes tanto para la comunidad en general como el sector infomediario. Además, los conjuntos de datos se presentan en diferentes formatos, complicando su reutilización. Para alcanzar los objetivos iniciales del proyecto es preciso una estrecha colaboración con el sector del I + D o el sector educativo porque tanto investigadores como estudiantes de ingeniería informática, mediante el desarrollo de servicios innovadores (aplicaciones web y móviles), están involucrados en la reutilización directa de datos.

Implementación

El software para montar un portal de datos puede ser desarrollado desde cero, adquiriendo una plataforma externa o usando software de código abierto. El ejemplo más conocido de un paquete de código abierto es CKAN.

Ejemplo de implementación

Según [este estudio](#) hay más de 1600 portales open data en el mundo.

Validación

El portal de datos abiertos debe cumplir determinadas funciones:

- Actuar como plataforma a través de la cual están disponibles los conjuntos de datos, su catalogación y permita buscarlos.
- Promover que se incluyan metadatos y facilitar añadirlos en el momento de la publicación.
- Actuar como escaparate para aplicaciones reutilizadoras de datos.
- Actuar como punto de encuentro para los diferentes actores de la comunidad open data.

Impacto



Reutilización de los datos



Acceso a los datos



Estrategia de publicación



Publicador

Crear de un portal de Gobierno Abierto para compartir datos (2/2)



Referencias

- 2017 Grupo Datos Abiertos de la FEMP [ES]: [Datos abiertos . Guía estratégica para su puesta en marcha. Conjuntos de datos mínimos a publicar](#)
- 2017 European data portal [EN]: [Extensive Study Recommendations for Open Data Portals: from setup to sustainability](#)
- 2015 datos.gob.es [ES]: [Guía para publicar datos abiertos de manera rápida y sencilla \(con CKAN\)](#)
- 2015 datos.gob.es [ES]: [Plataformas de publicación de datos abiertos](#)
- 2014 Open Data Monitor [EN]: [Open Data Topologies, Catalogues and Metadata Harmonization](#)
- 2014 Universidad de la Laguna [EN]: [Open Data to Improve Sharing and Publication of Information between Public Administrations](#)
- 2014 datos.gob.es [EN]: [A federation tool for opendata portals](#)
- 2013 Universidad de Trento, Trento Rise [EN]: [Open Government Data - Fostering Innovation](#)
- 2013 Agenzia per l'Italia Digitale [EN]: [Italian National Guidelines for the Valorization of the Public Sector Information](#)
- 2013 Athanasios Karalopoulos [EN]: [European Interoperability: The ISA Core Vocabularies](#)
- 2015 The European Data Portal [EN]: [The EDP: Architecture Overview](#)
- 2013 Fraunhofer-Institut für Offene Kommunikationssysteme [EN]: [Model-Driven Engineering for Data Harvesters](#)
- 2013 Philip Millard (Intrasoft), Jens Klessmann (Fraunhofer FOKUS) [EN]: [The Pan European Data Portal - Early Wireframes](#)
- 2013 The European Data Portal [EN]: [The European Data Portal - Opening up Europe's Public Data](#)
- 2013 The European Data Portal [EN]: [The EDP: A Technical View](#)
- 2013 The European Data Portal [EN]: [The Role of the Portal](#)



Poner datos a disposición a través de una API



Una API ofrece mayor flexibilidad y procesabilidad para los consumidores de los datos porque permite el uso de datos en tiempo real, el filtrado de peticiones y la capacidad de trabajar con los datos a nivel atómico. Seguramente una API sea la mejor opción para publicar cuando un conjunto de datos es grande, se actualiza con frecuencia o es muy complejo.



Estrategia de publicación

Reto a solucionar

Se pretende que los desarrolladores tengan acceso a los datos de manera programable para el uso en sus propias aplicaciones proporcionando datos actualizados sin requerir un esfuerzo adicional. Las aplicaciones web podrán obtener determinados datos desde una interfaz programable.

Implementación

La creación de una API es una tarea un poco más complicada que publicar datos para su descarga directa. Requiere una cierta comprensión de cómo construir una aplicación web. Sin embargo, no es precisamente necesario construirlas desde cero. Si se utiliza una plataforma de gestión de datos como CKAN es posible habilitar una API ya existente. Muchos marcos de desarrollo Web incluyen soporte para APIs, y también hay marcos diseñados específicamente para la creación de API personalizadas.

Rails (Ruby), Django (Python) y Express (NodeJS) son algunos ejemplos de marcos de desarrollo web que ofrecen soporte para la creación de APIs. Algunos ejemplos de marcos de API son Swagger, Apigility, Restify y Restlet.

Ejemplo de implementación

Además de proporcionar descargas masivas de datos sobre transporte público, se decide ofrecer un mecanismo de acceso a datos más flexible. Para ello, se desarrolla un API para ofrecer acceso a paradas de autobús, rutas de autobús e información en tiempo real sobre paradas de autobús. Ejemplos de su uso: <https://www.w3.org/TR/dwbp/dwbp-api-example.html>

Validación

Comprobar si un cliente puede simular llamadas de prueba y la API devuelve las respuestas esperadas.

Impacto



Reutilización de los datos



Legibilidad de los datos



Procesamiento de los datos

Referencias

- 2017 datos.gob.es [ES]: [Buenas prácticas en el diseño de APIs y Linked Data](#)
- 2017 European data portal [EN]: Extensive Study [Recommendations for Open Data Portals: from setup to sustainability](#)
- European Data Portal [EN]: Practical Guide [Technical preparation and implementation](#)
- European Data Portal [EN]: [Practical Guide Publishing data](#)
- 2015 W3C [EN]: [R-AccessRealTime](#)
- 2015 W3C [EN]: [R-AccessUpToDate](#)

Las APIs que se basan en estándares Web y aprovechan las fortalezas de la Web. Por ejemplo, el uso de verbos HTTP como métodos y URIs que se asignan directamente a recursos individuales ayudan a evitar el acoplamiento entre solicitudes y respuestas, lo que hace que una API sea fácil de mantener y pueda ser fácilmente entendida y utilizada por los desarrolladores. La descentralización de la Web es una fortaleza que permite el rápido crecimiento del uso de estas tecnologías.



Formatos



Estrategia de publicación

Reto a solucionar

Desarrollo de APIs basadas en estándares Web (como REST). La API debe ser más fácil de mantener.

Implementación

REST es un estilo arquitectónico que, cuando se utiliza en una API Web, aprovecha la arquitectura de la propia Web. Hay muchos entornos de desarrollo RESTful que permiten la implementación en cualquiera de los lenguajes de programación más usados. Otro aspecto de la implementación a considerar es la creación de una API hipermedia, que responde tanto a enlaces como a datos. Los enlaces son lo que hacen que la Web sea una red, y las API de datos pueden ser más útiles y utilizables mediante la inclusión de enlaces en sus respuestas. Los enlaces pueden ofrecer recursos adicionales, documentación y navegación. Incluso aunque una API no cumpla todas las restricciones de REST el devolver enlaces en las respuestas puede hacer que su servicio sea rico y auto-documentado.

Ejemplo de implementación

Un ejemplo de respuesta para la obtención de información sobre una determinada ruta de bus desde una API hipermedia podría tener el siguiente aspecto:

```
{
  "code": "200",
  "text": "OK",
  "data": {
    "update_time": "2013-01-01T03:00:02Z",
    "route_name": "Lexington South",
    "route_description": "Lexington corridor south of Market"
  },
  "links": [
    {
      "href": "http://data.mycity.example.com/transport/api/v2/routes/52",
      "rel": "self",
      "type": "application/json",
      "method": "GET"
    },
    {
      "method": "GET"
    }
  ]
}
```

[para más detalle, visitar <https://www.w3.org/TR/dwbp/#APIHttpVerbs>]

En cualquier llamada a la [API del catálogo de datos de datos.gob.es](http://datos.gob.es) se obtiene en los resultados el código de respuesta y los encabezados de la respuesta:

Código de la Respuesta: 200

Encabezados de la Respuesta

```
{
  "date": "Tue, 18 Jul 2017 09:53:39 GMT",
  "x-content-type-options": "nosniff",
  "age": "0",
  "x-frame-options": "SAMEORIGIN",
  "access-control-allow-methods": "POST, GET, OPTIONS, PUT, DELETE",
  "content-type": "application/json",
  "access-control-allow-origin": "*",
  "allow": "POST, GET, OPTIONS, PUT, DELETE",
  "transfer-encoding": "chunked",
  "accept-ranges": "bytes",
  "content-location": "http://datos.gob.es/apidata/catalog/dataset.json?_pageSize=10&_page=0&_sort=title",
  "vary": "Accept",
  "x-xss-protection": "1; mode=block"
}
```

Validación

Comprobar que el servicio evita usar HTTP como un túnel para llamadas a métodos personalizados y que los URI no contienen nombres de métodos.

Impacto



Interoperabilidad



Procesamiento de los datos



Acceso a los datos

Referencias

- 2017 datos.gob.es [ES]: [Buenas prácticas en el diseño de APIs y Linked Data](#)
- 2015 W3C [EN]: [R-API Documented](#)
- 2015 W3C [EN]: [R-Unique Identifier](#)



Publicador

Proporcionar documentación completa del API



Los desarrolladores son los principales consumidores de un API y la documentación es la primera impresión acerca de su calidad y utilidad. Cuando la documentación de la API es completa y fácil de entender, los desarrolladores probablemente están más dispuestos a continuar con su intención de usarla. Proporcionar una documentación completa permite a los desarrolladores crear código de manera más eficiente. Al resaltar los cambios, los usuarios pueden aprovechar las nuevas características y adaptar su código si es necesario.



Estrategia de publicación



Documentación/ Catalogación

Reto a solucionar

El objetivo que se plantea es que los desarrolladores sean capaces obtener información detallada sobre cada llamada a una API, incluyendo los parámetros que esta necesita y lo que se espera que devuelva, es decir, todo el conjunto de información relacionada con la API. El conjunto de valores, cómo usarla, avisos de cambios recientes, información de contacto, etc., deben describirse y obtenerse fácilmente.

Implementación

Una documentación típica de una API es proporcionar una lista completa de las llamadas que la API puede manejar, describiendo el propósito de cada una, detallando los parámetros que necesita y qué parámetros devuelve, dando uno o más ejemplos de su uso. Una buena tendencia en la documentación de la API es proporcionar un formulario en el que los desarrolladores puedan introducir llamadas específicas para realizar pruebas, para comprobar qué devuelve la API para cada caso de uso. Ahora hay herramientas disponibles para crear rápidamente este tipo de documentación, como Swagger, io-docs, OpenApis y otros. Es importante tener en cuenta que la API debe estar auto-documentada, es decir, que las llamadas deberían proporcionar información útil sobre errores y el uso de la API. Los usuarios deben poder ponerse en contacto con los administradores con preguntas, sugerencias o informes de errores. La calidad de la documentación también está relacionada con el uso y la retroalimentación de los desarrolladores. Es importante tratar de obtener comentarios constantes de los usuarios sobre la documentación de la API.

Ejemplo de implementación

Con el fin de ayudar a los desarrolladores, la agencia de transporte ofrece una [documentación](#) completa sobre la API que proporciona acceso a datos sobre paradas y rutas de autobús. La documentación de la API incluye una lista de las llamadas gestionadas por la API, los parámetros correspondientes y algunos ejemplos:

La Red de Consorcios de Transporte de Andalucía pone a disposición de los usuarios una documentación íntegra sobre el uso de su API, disponible en [este enlace](#). En ella podemos encontrar la configuración necesaria, una lista completa de todas las llamadas de las que dispone, así como todos los parámetros necesarios para poder realizarlas, detallando ejemplos de uso y de respuestas.

Validación

Comprobar si el conjunto de datos completo se puede recuperar con una sola solicitud.

Impacto



Reutilización de los datos



Legibilidad de los datos



Procesamiento de los datos

Referencias

- 2017 datos.gob.es [ES]: [Buenas prácticas en el diseño de APIs y Linked Data](#)
- 2015 W3C [EN]: [R-API/Documented](#)



Publicador

Evitar modificaciones sustanciales en el API (1/2)



Estrategia de publicación

Cuando los desarrolladores implementan un cliente basado en un API, pueden confiar en las características específicas que se han incorporado a dicha API, como el esquema o el formato de una respuesta. Evitar grandes cambios en el API minimiza los riesgos de rotura en el código del cliente. Comunicar los cambios que se realicen permite a los desarrolladores aprovechar las nuevas características y, en caso de una posible incompatibilidad, tomar medidas.

Reto a solucionar

El objetivo es que los desarrolladores sepan de las mejoras existentes y que sean capaces de hacer uso de ellas. Los cambios en las APIs no son frecuentes y, si ocurren, los desarrolladores tendrán suficiente tiempo e información para adaptar su código. Eso les permitirá evitar la discontinuación de compatibilidades, aumentando la confianza entre las partes. Los cambios en la API se anunciarán en el sitio de documentación de la API.

Implementación

Para mejorar la API, los esfuerzos se deben centrar en agregar nuevas llamadas o nuevas opciones en lugar de cambiar cómo funcionan las funcionalidades existentes. Quienes estén consumiendo esa API podrían ignorar esas nuevas funcionalidades y seguir funcionando sin problema.

Ejemplo de implementación

Algunos ejemplos de posibles cambios sustanciales en una API son:

- Eliminar una llamada
- Cambiar el método utilizado para realizar una llamada
- Cambiar el URI de un recurso utilizado en una llamada
- Agregar un parámetro necesario para una llamada
- Cambiar el tipo de datos de un parámetro
- Cambiar el nombre de una clave en una respuesta de valor-clave
- Cambiar la estructura de una respuesta XML
- Cambiar el tipo de datos de un valor en una respuesta, como cambiar una cadena a una matriz

Supongamos que la API de la agencia de transporte de MyCity responde a una solicitud de tiempo de llegada de un determinado autobús en una sola estación como `http://data.mycity.example.com/transport/api/arrivals/buses/53/stop/12`, pero la agencia decide que quiere hacer posible consultar una serie de paradas a la vez. En lugar de cambiar la forma de la solicitud para requerir un intervalo de paradas, como `http://data.mycity.example.com/transport/api/arrivals/buses/53/stop/12-12`, la agencia puede conservar la antigua llamada API y agregar una nueva para múltiples paradas, como `http://data.mycity.example.com/transport/api/arrivals/buses/53/stops/1-12`.

Validación

Se debe lanzar una versión de prueba de la API con la implementación de las nuevas características antes de aplicarlas a la versión de producción. Una opción es invitar a desarrolladores a probar sus aplicaciones con la nueva versión del API para que así se puedan recibir comentarios sobre ésta.



Publicador

Evitar modificaciones sustanciales en el API (1/2)



Impacto



Confianza entre las partes



Interoperabilidad



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [Buenas prácticas en el diseño de APIs y Linked Data](#)
- 2015 W3C [EN]: [R-PersistentIdentification](#)
- 2015 W3C [EN]: [R-APIDocumented](#)



Publicador

Proporcionar acceso en tiempo real



La presencia de datos en tiempo real en la Web permite el acceso a datos críticos sensibles al tiempo y fomenta el desarrollo de aplicaciones web en tiempo real. El acceso en tiempo real depende de los productores de datos que facilitan la disponibilidad de sus datos al editor de datos. La necesidad de proporcionar acceso en tiempo real para una aplicación determinada deberá evaluarse caso por caso considerando las frecuencias de actualización, la latencia introducida por los pasos posteriores al procesamiento de los datos, la disponibilidad de la infraestructura y los datos que necesitan los consumidores. Además de facilitar la accesibilidad de los datos, los editores pueden proporcionar información adicional que describa dichos errores, anomalías y retrasos en la publicación.



Estrategia de publicación

Reto a solucionar

El objetivo es que las aplicaciones puedan acceder a datos críticos en tiempo real o casi en tiempo real, entendiéndose por tiempo real un intervalo de milisegundos a unos pocos segundos después de la creación de los datos.

Implementación

Un posible enfoque para la implementación es la configuración de un servicio web que proporcione una conexión para que los datos en tiempo real sean recibidos y puestos a disposición de los consumidores de manera instantánea mediante polling o streaming.

Si los datos no son consultados frecuentemente por los consumidores, los datos en tiempo real pueden ser servidos a petición del consumidor a través de una API proporcionada por el organismo publicador.

Si los datos son consultados frecuentemente, un servicio de datos en streaming puede ser la opción más apropiada, enviando los datos a través de una API. Si bien las técnicas de streaming están fuera del alcance de esta práctica recomendada, hay muchos protocolos y tecnologías estándar disponibles (por ejemplo WebSocket o EventSourceAPI) para conseguir que los clientes reciban actualizaciones automáticas desde el servidor.

Ejemplo de implementación

La Agencia de Transporte de MyCity realiza un seguimiento de todos los datos GPS del bus. La API proporciona a los consumidores información de estado en tiempo real utilizando una API REST. La API permite al consumidor seleccionar:

- Posición actual del autobús
- Hora de llegada del autobús
- Estado del bus

La aplicación [Incidentes del Tráfico](#) facilitada por la iniciativa de datos abiertos del Ayuntamiento de Lorca, proporciona información en tiempo real sobre todas las incidencias del tráfico que suceden en la ciudad de Lorca, tales como calles cortadas, estrechamientos de calzada, reserva de espacio... georreferenciadas en un mapa para facilitar la localización de dicha incidencia.

Validación

Para probar adecuadamente el acceso a datos en tiempo real, los datos tendrán que ser rastreados desde el momento en que son recogidos hasta el momento de su publicación y acceso.

Impacto



Reutilización de los datos



Acceso a los datos

Referencias

- 2015 W3C [EN]: [R-AccessRealTime](#)



Publicador

Proporcionar datos a coste cero



Promover un servicio RISP sin ningún tipo de coste ayuda a alcanzar un máximo potencial comercial y no comercial. Para la reutilización comercial, la carga cero para el PSI significa que el desarrollo de productos y servicios tendrá un costo menor, aumentando así su viabilidad. Para la reutilización no comercial, los el coste cero facilita la reutilización que de otro modo no sería posible.



Negocio



Estrategia de publicación

Reto a solucionar

Se ha observado que el sector público tiene dificultades para determinar el potencial comercial y no comercial de la RISP. También es cierto que al cobrar por la reutilización de la información del sector público, el potencial de dicha reutilización se verá reducido.

Implementación

En la medida de lo posible, las organizaciones del sector público deberían publicar una estrategia RISP a coste cero para reducir las barreras para su reutilización.

Ejemplo de implementación

Desarrollar e implementar una política que aliente a las organizaciones del sector público a reducir el costo marginal y proporcionar acceso a los datos a coste cero. En ocasiones puede ser necesario un financiamiento adicional para las agencias que proveen RISP a coste cero.

La Junta de Castilla y León publica en su portal [Datos Abiertos de Castilla y León](#) información de forma libre (sin restricciones de copyright, patentes u otros mecanismos de control) para que sea reutilizada, permitiendo que cualquier persona u organización pueda construir sobre ellos una nueva idea que resulte en nuevos datos, conocimientos, mejorar procesos, dar valor añadido a los existentes o incluso crear nuevos servicios.

Impacto



Reutilización de los datos



Acceso a los datos

Referencias

- 2017 datos.gob.es [ES]: [Tendencias y buenas prácticas en la implementación de políticas de datos abiertos](#)
- 2017 datos.gob.es [ES]: [Conceptos básicos, beneficios del open data y barreras](#)
- 2013 datos.gob.es [ES]: [Informe: RISP - Gratuidad, tasas y precios públicos](#)
- 2013 Lorenzo Modena, CEO OpenMo [EN]: [How Trentino opened public transportation data and benefitted of a mobile ticketing solution for free](#)
- 2016 The Open Group: Business Scenario [EN]: [Open Public Sector Data – Detailed Objectives](#)
- 2013 Mª Dolores Hernández Maroto [EN]: [A Federation Tool for Open Data Portals](#)



Proporcionar presentación de la información complementaria



Los datos publicados deberían llegar a todos los interesados, pero si únicamente está disponible un conjunto de datos mediante API o descarga directa, queda de la mano de los consumidores la tarea de interpretación. La Web ofrece un gran abanico de oportunidades para representar datos de manera que se permita a los usuarios explorarlos y aprender sin tener que crear sus propias herramientas.



Estrategia de publicación

Reto a solucionar

Se pretende que las representaciones adicionales de datos permitan a los usuarios tener una visión inmediata de los datos de una manera en que se entiendan fácilmente.

Implementación

Una forma muy sencilla de proporcionar una visión inmediata de los datos es publicar un resumen analítico en una página HTML. Incluir datos en forma gráficos o tablas puede ayudar a los usuarios comprender rápidamente el significado de los datos. Si se dispone de los medios para crear visualizaciones interactivas o aplicaciones Web que usen los datos, se puede dar a los consumidores una mayor capacidad para comprenderlos y descubrir patrones en ellos. Estas representaciones adicionales también demuestran que los datos se pueden procesar y fomentan su reutilización.

Ejemplo de implementación

La agencia de transporte de MyCity publica datos detallados sobre todas sus líneas de tránsito a través de una API, pero también tiene muchos usuarios que no son desarrolladores web y que quieren saber cómo usar este sistema para moverse por la ciudad. La agencia de tránsito podría construir una aplicación web que permitiese a los usuarios introducir una dirección de salida y un destino y recibir instrucciones paso a paso para hacer su viaje a través del transporte público.

Validación

Comprobar que el conjunto de datos esté acompañado de algún contenido interpretativo adicional que pueda percibirse sin descargar los datos o invocar una API.

Impacto



Legibilidad de los datos



Estrategia de publicación

Referencias

- 2017 European data portal [EN]: Extensive Study [Recommendations for Open Data Portals: from setup to sustainability](#)
- European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2015 W3C [EN]: [R-DataEnrichment](#)



Publicador

Proporcionar la posibilidad de una descarga completa de datos



Cuando los datos se distribuyen a través de muchos URIs y pueden organizarse como un solo contenedor, el acceso a los datos en bruto puede resultar útil. El acceso masivo proporciona un medio consistente para manejar los datos como un conjunto de datos. El acceso individual a datos puede resultar engorroso y, si se utiliza para reensamblar el conjunto de datos completo, puede conducir a enfoques inconsistentes para manejar los datos.



Estrategia de publicación

Reto a solucionar

El acceso individual a datos puede resultar engorroso y, si se utiliza para reensamblar el conjunto de datos completo, puede conducir a enfoques inconsistentes para manejar los datos. Debe tenerse en cuenta que la transferencia de ficheros grandes suele llevar más tiempo del que los usuarios consideran razonable. Estudiar la posibilidad de usar protocolos específicos para la transferencia de ficheros.

Implementación

Dependiendo de la naturaleza de los datos y las necesidades de los consumidores, podrían existir los siguientes enfoques:

- Para conjuntos de datos que inicialmente existen como archivos múltiples, de debe preprocesar una copia de los datos en un solo archivo y hacer que los datos sean accesibles para su descarga desde un URI. Para conjuntos de datos más grandes, el archivo también se puede comprimir.
- El proporcionar una API que incluya la capacidad de realizar una descarga masiva además de consultas dinámicas.
- Para conjuntos de datos muy grandes, las transferencias de archivos masivos se pueden realizar a través de medios distintos a http, como bbcp o GridFTP.

La descarga masiva debe incluir los metadatos que describen el conjunto de datos.

Ejemplo de implementación

La agencia de transporte de MyCity tiene un gran conjunto de datos con tiempos de llegada de los diversos medios de transporte que se obtuvieron a lo largo del año. Los datos pueden almacenarse mensualmente como un archivo CSV. Supongamos que la agencia quiere hacer que, para un hackathon, los datos estén disponibles mediante una descarga masiva que contenga todos los archivos CSV. Dado que todos los datos de llegada de todos los servicios de tránsito serían una gran cantidad de datos, y que quieren proporcionar todos los meses juntos como un conjunto de datos, se opta por ofrecer un único archivo comprimido que contenga todos los datos.

Validación

Comprobar si el conjunto de datos completo se puede recuperar con una sola solicitud.

Impacto



Reutilización de los datos



Procesamiento de los datos



Acceso a los datos

Referencias

- European data portal [EN]: Data formats: [Most common data formats used](#)
- 2015 W3C [EN]: [R-AccessBulk](#)



Proporcionar una explicación para los datos que no están disponibles



La publicación online de documentación sobre datos que no se encuentran disponibles permite a los publicadores identificar explícitamente las lagunas de conocimiento existentes. Para los consumidores de los datos, esto proporciona una explicación del contexto y fomenta el uso de los datos disponibles.



Estrategia de publicación



Documentación/
Catalogación

Reto a solucionar

Se trata de conseguir que los consumidores sepan que los datos a los que se hace referencia desde el conjunto de datos actual no están disponibles o sólo están disponibles bajo ciertas condiciones.

Implementación

Dependiendo del contexto hay una varias maneras de indicar indisponibilidad de datos. Los editores de datos pueden publicar un documento HTML que proporcione una explicación legible para esta no disponibilidad de los datos. Desde la perspectiva de una interfaz de aplicación, se pueden usar códigos de estado HTTP apropiados con mensajes personalizables legibles por el usuario. Ejemplos de códigos de estado incluyen: (see other), 410 (permanently removed), 503 (service *providing data* unavailable).

Ejemplo de implementación

El conjunto de datos creado para las paradas de bus puede contener datos confidenciales, por ejemplo, información sobre el controlador de bus. En este caso, el editor proporciona una explicación que informa a los usuarios potenciales de que los datos personales sobre el controlador del bus no están disponibles.

Validación

Cuando el conjunto de datos incluye referencias a datos que ya no están disponibles o no están disponibles para todos los usuarios, se debe comprobar que se da una explicación e instrucciones para obtener el acceso a dichos datos(si es posible). Comprobar también si se devuelve un código de respuesta HTTP legítimo en el rango 400 o 500 al intentar obtener datos no disponibles.

Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2015 W3C [EN]: [R-AccessLevel](#)
- 2015 W3C [EN]: [R-SensitivePrivacy](#)
- 2015 W3C [EN]: [R-SensitiveSecurity](#)

MEDIR Y EVALUAR



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta



Métricas holísticas



En general, la evaluación de los beneficios de las acciones exclusivamente a escala local llevará a micro-optimizaciones y a la pérdida de oportunidades en un contexto más amplio. Esto no sólo es desventajoso en el caso de decidir acerca de la apertura de datos e información, sino que es un remanente administrativo desde tiempos en los que la medida holística era imposible debido a la inexistencia de sistemas de información integrados que pudiesen proporcionar una visión general rápida y completa sobre la formulación de políticas.



Estrategia de publicación



Negocio

Reto a solucionar

La publicación de datos e información de acuerdo con regulaciones, principios, buenas prácticas o recomendaciones generalmente tiene un efecto positivo. Sin embargo, en los casos en que el editor obtiene un ingreso sustancial mediante la monetización de los datos, la pérdida de dichos ingresos puede contribuir a tomar decisiones en contra de la publicación de datos o información. Por lo tanto, se requieren evaluaciones más sofisticadas que sean adecuadas para justificar la pérdida de ingresos en cierto nivel a cambio de obtener beneficios globales.

Implementación

Una serie de medidas tomadas a un nivel más alto a menudo mostrarán beneficios significativos para la organización como un todo, dando lugar a una mayor eficiencia en general, mejores servicios y una mayor transparencia.

Ejemplo de implementación

Algunos ejemplos de implementación de esta BP son:

- Austria: [Wirkungsorientierte Steuerung](#)
- Reino Unido: [Performance UK](#)
- Finlandia: [Government's analysis, assessment and research activities](#)
- Italia: [Impact of open data measured by indicators](#)

Impacto



Estrategia de publicación

Referencias

- 2014 datos.gob.es [ES]: [Informe de Tendencias e Iniciativas de Datos Abiertos](#)
- 2013 Heather Broomfield, Steinar Skagemo [EN]: [The Potential within the Government for Innovation and Efficiency from Open Data – Examples from the Norwegian public Sector](#)
- 2015 Heli Koski [EN]: [The Impacts of Open Data: Towards Ex Post Assessment](#)
- Open Data Institute method report [EN]: [Assessment tools for open data initiatives](#)
- Open Data Institute technical report [EN]: [Benchmarking open data automatically](#)



Publicador Reutilizador

Evaluación de la calidad de los datos abiertos (1/2)



La calidad de los datos se percibe principalmente como un término subjetivo: lo que es suficientemente bueno para unos puede no serlo para otros. Sin embargo, además del aspecto subjetivo, hay una visión objetiva que puede medirse y ayudar a establecer métricas comprensibles en términos de calidad del dato. La adhesión a los estándares, reforzados por herramientas que están incorporadas y utilizadas por los procesos, ayudará a aumentar la calidad de los datos. Para mantener un incremento de la calidad es preciso medirla en todos los estados por los que pasa el dato y no solo en la fase de publicación.



Reto a solucionar

La proliferación de datos abiertos ha supuesto el fomento de los procesos de innovación para la mejora o creación de nuevos productos o servicios, el aumento de la transparencia y la evaluación de las políticas públicas. Pero también ha planteado inquietudes sobre la calidad de los recursos proporcionados. El supuesto inicial de que más datos, incluso de calidad y origen inciertos, darán incondicionalmente como resultado mejores decisiones llevó a la conclusión de que el principio "garbage-in, garbage-out" (basura entra, basura sale) sigue siendo válido. Este hecho combinado con el aumento de las preocupaciones sobre la usabilidad de la plataforma de datos, la alfabetización de datos y la confianza han hecho poner el foco en el aspecto de la calidad.

Implementación

La implementación requiere afrontar el problema tanto desde una perspectiva técnica como organizativa:

- Técnica: utilización de normas y estándares. El CIO debe proporcionar orientación sobre las normas y estándares a utilizar. Es responsabilidad suya identificar entidades de datos semánticamente equivalentes para utilizar vocabularios comunes.
- Organizativa: el CIO debe implementar un marco de gestión de datos y metadatos.

Ejemplo de implementación

Algunos ejemplos de la implementación de esta BP son:

- Austria Mission Statement of the Sub-working Group [Quality Assurance of Open Data Portals](#) of the Cooperation Open Government Data Austria
- UK [Cross platform character encoding profile](#)
- UK [ODI Certificate for the Westminster City Council](#)
- Serbia [Validating RDF Data Cube Models](#)
- Finlandia [Valmistele ja avaa - Prepare and open](#) Section 3.6. Tiedon viimeistely ja laatu - Finishing the data and data quality

Validación

Las métricas tradicionales para evaluar la calidad de los datos, como la precisión, la aplicabilidad y la comprensión, siguen siendo pertinentes y, en el ámbito de los datos abiertos, se amplían mediante medidas como la apertura, la oportunidad y la primacía. El trabajo realizado en el proyecto Open Data Support de la Comisión Europea sugiere nueve aspectos a considerar:

- Precisión: ¿Los datos que se corresponden correctamente con el ente o evento del mundo real?
- Consistencia: ¿Los datos contienen contradicciones?
- Disponibilidad: ¿Se puede acceder a los datos ahora y lo largo el tiempo?
- Integridad: ¿Incluyen los datos todos los elementos de datos que representan la entidad o el evento?
- Conformidad: ¿Los datos siguen los estándares aceptados?
- Credibilidad: ¿Los datos están basados en fuentes fiables?
- Procesabilidad: ¿Los datos son legibles por máquinas?
- Relevancia: ¿Los datos incluyen una cantidad apropiada de datos?
- Oportunidad: ¿Los datos representan la situación real y se publican lo suficientemente pronto?

Impacto



Confianza entre las partes



Interoperabilidad



Estrategia de publicación

Referencias

- 2017 datos.gob.es [ES]: [Pautas metodológicas para la apertura de datos](#)
- 2017 European data portal [EN]: [Extensive Study Recommendations for Open Data Portals: from setup to sustainability](#)
- 2017 European data portal [EN]: [Analytical Report Barriers in working with Open Data](#)
- 2017 European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2012 David Corsar, Peter Edwards (Universidad de Aberdeen) [EN]: [Enhancing Open Data with Provenance](#)
- 2014 ProvenanceWeek [EN]: [ProvenanceWeek 2014](#)
- 2014 Giorgos Flouris, Yannis Roussakis, Marria Poveda-Villalon, Pablo N. Mendes, Irini Fundulaki [EN]: [Using Provenance for Quality Assessment and Repair in Linked Open Data](#)
- 2015 Makx Dekkers (AMI Consult) [EN]: [How good is good enough?](#)
- 2013 Amanda Smith, Sumika Sakanishi [EN]: [Publishing and improving the quality of open data with Open Data Certificates](#)
- 2014 Samos presentation [EN]: [Examples from the Norwegian public sector](#)
- 2014 Lisbon workshop session [EN]: [Roadblocks in Commercial Open Data Usage](#)
- 2015 Leigh Dodds [EN]: [Comparing the 5-star scheme with Open Data Certificates](#)



Publicador

Proporcionar información sobre la calidad de los datos (1/2)



La calidad de los datos podría afectar sensiblemente a la idoneidad de estos para ser usados por aplicaciones específicas, incluidas aplicaciones muy diferentes al propósito para el que los datos se generaron originalmente. La documentación de la calidad de los datos facilita considerablemente el proceso de selección de conjuntos de datos, aumentando las posibilidades de reutilización. Independientemente de las peculiaridades específicas del dominio, la calidad de los datos debe ser documentada y los problemas de calidad conocidos deben ser expresados explícitamente en los metadatos.



Reto a solucionar

Hacer que los seres humanos y los agentes software puedan evaluar la calidad y por lo tanto la idoneidad de un conjunto de datos para una aplicación concreta.

Implementación

La versión legible por máquinas de los metadatos de calidad de datos puede proporcionarse utilizando el vocabulario de calidad de datos desarrollado por el grupo de trabajo DWBP [VOCAB-DQV].

Ejemplo de implementación

Legibilidad por humanos:
En la siguiente [página de ejemplo](#) se muestra información de calidad de datos legible por humanos

Legibilidad por máquinas:
El siguiente ejemplo muestra los metadatos para una distribución en CSV del dataset de paradas de bus con la inclusión de los metadatos de calidad de datos. Los metadatos se definieron de acuerdo con el vocabulario de calidad de datos:

```

:stops-2015-05-05.csv
  a dcat:Distribution ;
  dcat:downloadURL <http://data.mycity.example.com/transport/dataset/bus/stops-2015-05-05.csv> ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8" ;
  dct:license <http://creativecommons.org/licenses/by-sa/3.0/> ;
  dqv:hasQualityMeasurement :measure1, :measure2
.
:measure1
  a dqv:QualityMeasurement ;
  dqv:computedOn :stops-2015-05-05.csv ;
  dqv:isMeasurementOf :downloadURLAvailabilityMetric ;
  dqv:value "true"^^xsd:boolean

```

Validación

Comprobar que los metadatos del conjunto de datos incluyen información de calidad sobre el conjunto de datos.
Comprobar si una aplicación informática puede procesar automáticamente la información de calidad sobre el conjunto de datos.



Publicador

Proporcionar información sobre la calidad de los datos (2/2)



Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2017 European Data Portal [EN]: Practical Guide [Preparing data](#)
- 2017 European data portal: Analytical Report [EN]: [Barriers in working with Open Data](#)
- 2015 W3C [EN]: [R-QualityMetrics](#)
- 2015 W3C [EN]: [R-DataMissingIncomplete](#)
- 2015 W3C [EN]: [R-QualityOpinions](#)



Habilitar canales de feedback para mejorar la calidad de los datos gubernamentales existentes



El objetivo de esta buena práctica es mejorar la calidad de los datos de la administración con el feedback de usuarios: aviso de errores, incoherencias, datos incompletos... Esto está en consonancia con la Comunicación de la Comisión Europea 2014/C240/01, párrafo 3.2, que dice: "Para facilitar el uso de datos en el sector público y aumentar significativamente el valor de los conjuntos de datos para su posterior reutilización, se recomienda que los conjuntos de datos estén [...] sujetos a la retroalimentación periódica de los usuarios (consulta pública, Blogs, informes automatizados, etc.) para mantener la calidad en el tiempo y promover la participación pública."



Estrategia de publicación



Comunidad

Reto a solucionar

A menudo, los reutilizadores de conjuntos de datos gubernamentales realizan copias de estos datos con el fin de, por ejemplo, corregir errores o completar datos. Por un lado, este enfoque no es óptimo, ya que conduce a la duplicación de esfuerzos y reduce la posibilidad de compartir y reutilizar, y por otro lado, el publicador quiere tener información sobre las necesidades de los usuarios y los beneficios de abrir los datos. Por lo tanto, la institución que publica los datos puede plantear las siguientes preguntas:

- ¿Quién está accediendo y usando mis datos? ¿Los datos satisfacen su necesidad?
- ¿Cuáles son sus experiencias?
- ¿Cuál es el verdadero valor de mis datos?
- ¿Podemos aumentar nuestros ingresos proporcionando mejores servicios a nuestros clientes?

Implementación

Los organismos del sector público responsables de publicar datos deben proporcionar mecanismos de feedback a través de los cuales las partes interesadas puedan identificar errores y corregirlos cuando sea posible. Un posible enfoque práctico podría ser utilizar un sistema distribuido de control de versiones para los datos publicados, como GitHub común para el software de código abierto. Alternativamente, se puede proporcionar un simple canal retroalimentación utilizando cajas de comentarios, foros, etc. Los organismos del sector público deberían alentar activamente a los usuarios interesados a usar estos mecanismos para recopilar comentarios de los usuarios y mejorar la calidad de las fuentes de datos gubernamentales existentes.

Ejemplo de implementación

Algunos ejemplos de la implementación de esta práctica son:

- Polonia - [Implementation](#) of the revised re-use Directive in Poland, Open Data Portal
- Escocia - [ALISS](#) service
- Unión Europea - FP7 [ENGAGE](#) project feedback mechanisms based on Web 2.0
- Republica Checa - [Fórum pro otevřená data](#)

Impacto



Confianza entre las partes

Referencias

- 2016 European data portal [EN]: Reports about Open Data [Co-creating Public Policies or Ways to Bring Citizens into the Process](#)
- 2014 Mateja Prešern, Ministry of Public Administration [EN]: [Supervisor application, Re-use of PSI with a goal of strengthening the integrity and transparency](#)
- 2013 Lorenzo Canova, Antonio Vetrò, Marco Torchiano, Raimondo lemma & Federico Morando, Politecnico di Torino [EN]: [OpenCoesion and Monithon - a Transparency Effort](#)
- 2015 Jacek Wolszczak, Ministry of Administration and Digitization [EN]: [Implementation of the revised re-use Directive in Poland](#),
- 2015 Amanda Smith & Sumika Sakanishi, ODI [EN]: [Publishing and improving the quality of open data with Open Data Certificates](#)
- 2014 Charalampos Alexopoulos, Euripidis Loukis, Yannis Charalabidis [EN]: [A Platform for Closing the Open Data Feedback Loop based on Web 2.0 functionality](#)



Publicador

Disponer de un sistema de feedback (1/2)



Al compartir feedback con los consumidores, los editores pueden demostrar a los usuarios que sus preocupaciones están siendo atendidas. Esta práctica también ayuda a los consumidores a entender cualquier problema que pueda afectar a su capacidad de usar los datos y puede fomentar un sentimiento de comunidad entre ellos.



Comunidad

Reto a solucionar

Se trata de conseguir que los consumidores puedan evaluar los tipos de errores que afectan a los conjuntos de datos, revisar las experiencias de otros usuarios y asegurarse de que el editor está abordando de forma activa los problemas que surgen. Los consumidores también podrán determinar si otros usuarios ya han proporcionado comentarios similares, ahorrándoles la molestia de presentar informes de errores innecesarios.

Implementación

Puede habilitarse o incorporarse el feedback como parte de una página Web HTML, pero también puede proporcionarse en un formato legible por máquina usando el Dataset Usage Vocabulary.

Ejemplo de implementación

```

:stops-2015-05-05
  a dcat:Dataset ;
  dct:title "Bus stops of MyCity" ;
  dcat:keyword "transport","mobility","bus" ;
  dct:issued "2015-05-05"^^xsd:date ;
  dcat:contactPoint <http://data.mycity.example.com/transport/contact> ;
  dct:temporal <http://reference.data.gov.uk/id/year/2015> ;
  dct:spatial <http://sws.geonames.org/3399415> ;
  dct:publisher :transport-agency-mycity ;
  dct:accrualPeriodicity <http://purl.org/linked-data/sdmx/2009/code#freq-A> ;
  dcat:theme :mobility ;
  dcat:distribution :stops-2015-05-05.csv
.

:stops-2015-05-05.csv
  a dcat:Distribution ;
  dct:title "CSV distribution of stops-2015-05-05 dataset" ;
  dct:description "CSV distribution of the bus stops dataset of MyCity" ;
  dcat:mediaType "text/csv;charset=UTF-8"
.

:comment1Content
  a oa:TextualBody ;
  rdf:value "This dataset is missing stop 3"
.

:comment1
  a oa:Annotation ;
  a duv:UserFeedback ;
  oa:hasBody :comment1Content ;
  oa:hasTarget :stops-2015-05-05 ;
  dct:creator :localresident ;
  oa:motivatedBy oa:assessing
.

```

para más detalle, visitar <https://www.w3.org/TR/dwbp/#StructuralMetadata>



Publicador

Disponer de un sistema de feedback (2/2)



Validación

Comprobar que cualquier feedback dado por consumidores de datos para un conjunto de datos específico o distribución esté disponible públicamente.

Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2017 European data portal [EN]: Extensive Study [Recommendations for Open Data Portals: from setup to sustainability](#)
- 2016 European data portal [EN]: Reports about Open Data [Co-creating Public Policies or Ways to Bring Citizens into the Process](#)
- 2015 W3C [EN]: [R-UsageFeedback](#)
- 2015 W3C [EN]: [R-QualityOpinions](#)



Publicador

Obtención de feedback de los consumidores de datos



La obtención de feedback ayuda a los editores a comprender las necesidades de los consumidores de datos y puede ayudarlos a mejorar la calidad de los datos que publican. También mejora la confianza mostrando a los consumidores que el editor se preocupa por atender sus necesidades. Especificar claramente un mecanismo de retroalimentación elimina la barrera de tener que buscar una forma de proporcionar retroalimentación.



Comunidad

Reto a solucionar

El objetivo es que los consumidores de datos puedan proporcionar retroalimentación y calificaciones sobre conjuntos de datos y distribuciones.

Implementación

Se debe proporcionar a los consumidores de datos uno o más mecanismos de retroalimentación que incluyan, pero que no se limiten a un formulario de contacto, sistema de puntuación y haga clic en los botones de calificación de calidad de datos o un cuadro de comentarios. Con el fin de aprovechar al máximo la retroalimentación recibida de los consumidores, es una buena idea recoger la retroalimentación con un sistema de seguimiento que almacene cada elemento en una base de datos, lo que permitiría su análisis y cuantificación. También es una buena idea capturar el tipo de cada elemento de retroalimentación, es decir, su motivación (edición, puntuaciones, comentarios o preguntas) de manera que cada elemento pueda expresarse usando el Dataset Usage Vocabulary.

Ejemplo de implementación

En la siguiente [página de ejemplo](#) se muestra un ejemplo de un formulario para la obtención de feedback.

En todos los datasets proporcionados por el [Portal de datos abiertos del Ayuntamiento de Madrid](#) existe la sección "Vota este contenido" donde se puede puntuar el contenido del dataset con una puntuación de entre 1 y 5. También existe una caja de comentarios donde el usuario puede dejar por escrito su opinión y dudas sobre el dataset en cuestión.

Validación

Comprobar que se proporciona al menos un mecanismo de retroalimentación y que es fácilmente detectable por los consumidores de datos.

Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- 2017 datos.gov.es [ES]: [Tendencias y buenas prácticas en la implementación de políticas de datos abiertos](#)
- 2016 European data portal [EN]: Reports about Open Data [Co-creating Public Policies or Ways to Bring Citizens into the Process](#)
- W3C [EN]: [R-UsageFeedback](#)
- W3C [EN]: [R-QualityOpinions](#)



Reutilizador

Proporcionar feedback al publicador original (1/2)



Los publicadores de los datos generalmente quieren saber si los datos que publican son útiles y además, puede que tengan que reportar estadísticas de uso para asignar recursos a las actividades de publicación de datos. Dar información sobre el uso de los datos les ayuda a justificar el esfuerzo de liberar los datos. Proporcionar feedback repercute en los publicadores, ayudándoles a mejorar sus datos para futuros usuarios.



Calidad del dato



Comunidad

Reto a solucionar

El objetivo es proporcionar una mejor comunicación que facilite a los publicadores determinar cómo se están utilizando los datos que publican, lo que a su vez les ayuda a justificar la publicación de los datos. También que los publicadores recibirán información de los pasos que pueden realizar para mejorar sus datos. Esto conduce a más y mejores datos para todos.

Implementación

Cuando empiece a utilizar un nuevo conjunto de datos, se debe tomar nota de la información de contacto del publicador de los conjuntos de datos que utiliza y de la fecha en que se puso en contacto con ellos. Debe seguirse la ruta indicada por el publicador para proporcionar comentarios. Si no proporcionan una ruta, se buscará información de contacto en el sitio web que aloja los datos.

Ejemplo de implementación

Cuando comience a utilizar un nuevo conjunto de datos, tome nota de la información de contacto del editor, la URI del conjunto de datos que utilizó y la fecha en la que se puso en contacto con ellos. Esto se puede hacer en comentarios dentro de su código donde se utiliza el dataset. Siga la ruta preferida del editor para proporcionar comentarios. Si no proporcionan una ruta, busque información de contacto para el sitio Web que aloja los datos.

Llamar a la API de transporte de MyCity, <http://data.mycity.example.com/transport/api/docs/>
Publicado por MyCity Transit Agency,
Se notificó de nuestra reutilización por correo electrónico a opendata@mycitytransit.example.org

En el dataset [Calidad del aire: Datos en tiempo real](#), publicado en el Portal de datos abiertos del Ayuntamiento de Madrid, vemos que se ha obtenido una cantidad considerable de votos, más de 3300, puntuando el contenido del dataset con una nota media de 2 sobre 5. También se han obtenido numerosos comentarios reflejando las dudas y críticas al contenido del dataset por parte de los usuarios. Se observa que el administrador de datos abiertos del portal interactúa con los usuarios intentando resolver las dudas e incidencias relativas a los comentarios proporcionados por ellos

Validación

Comprobar que al menos hay una comunicación de un usuario informando al publicador sobre el uso de dichos datos.



Reutilizador

Proporcionar feedback al publicador original (2/2)



Impacto



Reutilización de los datos



Confianza entre las partes

Referencias

- European data portal [EN]: Reports about Open Data [Co-creating Public Policies or Ways to Bring Citizens into the Process](#)
- 2015 W3C [EN]: [R-TrackDataUsage](#)
- 2015 W3C [EN]: [R-UsageFeedback](#)
- 2015 W3C [EN]: [R-QualityOpinions](#)

Iniciativa
aporta



GOBIERNO
DE ESPAÑA

MINISTERIO
DE HACIENDA
Y FUNCIÓN PÚBLICA

MINISTERIO
DE ENERGÍA, TURISMO
Y AGENDA DIGITAL

red.es

Iniciativa
aporta