



Universitat d'Alacant
Universidad de Alicante

Modelo de madurez para portales de
datos abiertos e incorporación a la
norma técnica nacional de Costa Rica

Edgar Oviedo Blanco



Tesis

Doctorales

www.eltallerdigital.com

UNIVERSIDAD de ALICANTE



Universitat d'Alacant
Universidad de Alicante

Departamento de Lenguajes y Sistemas Informáticos

Tesis Doctoral

Programa de Doctorado en Aplicaciones de la Informática

**Modelo de madurez para portales de datos abiertos e
incorporación a la norma técnica nacional de Costa Rica**

Autor:

Ing. Edgar Oviedo Blanco

Directores:

Dr. Jose Norberto Mazón

Dr. Jose Jacobo Zubcoff

Alicante, España

Enero 2016



Universitat d'Alacant
Universidad de Alicante

Programa de Doctorado en Aplicaciones de la Informática

Tesis Doctoral
Ing. Edgar Oviedo Blanco



Universitat d'Alacant
Universidad de Alicante



Agradecimientos

A Dios,
por la vida.

A Ana Cristina, Fabián, Felipe, Arianna y Lucía,
por ese apoyo incondicional de siempre
que ha sido vital para poder finalizar mis estudios doctorales.

Al Ministerio de Ciencia, Tecnología y Telecomunicaciones de Costa Rica (MICITT)
y al Consejo Nacional para Investigaciones Científicas y Tecnológicas (CONICIT),
por la confianza y el apoyo a lo largo de mis estudios de posgrado.

A mis directores, Jose Norberto Mazón y Jose Jacobo Zubcoff,
por su formidable apoyo en el desarrollo de esta investigación.

A INTECO,
por la confianza brindada al tomar en cuenta los resultados de mi investigación
para el desarrollo de la Norma Técnica para Portales de Datos Abiertos en Costa Rica.



Universitat d'Alacant
Universidad de Alicante

Programa de Doctorado en Aplicaciones de la Informática

Tesis Doctoral
Ing. Edgar Oviedo Blanco



Universitat d'Alacant
Universidad de Alicante



Prefacio

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

El crecimiento exponencial del volumen de datos disponibles hoy en día, ha generado la necesidad de integrar y clasificar dichos datos, posibilitando su publicación de tal manera que se facilite su posterior consumo. En este escenario de datos masivos (*"big data"*), tener en cuenta esta dicotomía publicación-consumo es de gran importancia para poder obtener valor económico y social a partir de la reutilización de datos. De hecho, ésta es la clave del concepto de datos abiertos (*"open data"*), que consiste en publicar datos de manera organizada con el fin de potenciar su consumo.

La publicación de estos datos se lleva a cabo mediante el desarrollo de portales de datos abiertos, con el fin de propiciar que puedan estar disponibles con un nivel de calidad adecuado para los consumidores de datos. Cuando estos datos abiertos provienen del sector público, se emplea el término datos públicos (*"public data"*), concepto que está relacionado directamente con la información de carácter gubernamental generada por instituciones públicas, puesta a disposición de la ciudadanía con el afán de fortalecer principalmente la transparencia y la rendición de cuentas, dos conceptos que – políticamente hablando – tienen alta relevancia en los gobiernos actuales. Además del impacto social que ello representa, también es posible obtener de los datos públicos un impacto económico si se cumple con ciertas características o criterios de calidad que aseguren la madurez del portal de datos abiertos en el que están publicados, garantizando que se permita una adecuada reutilización de los datos con el fin de obtener productos y servicios de valor añadido. Queda patente entonces la importancia de definir mecanismos adecuados para evaluar la madurez de un portal de datos abiertos.

Por tanto, la hipótesis de partida de la investigación desarrollada en esta tesis doctoral plantea que, a partir de ciertos criterios de calidad del portal de datos abiertos se puede determinar su madurez, lo que permitirá determinar el cumplimiento de sus objetivos. Estos criterios de calidad de un portal de datos abiertos deben contemplar aspectos tanto de la calidad de los datos como de la calidad del portal en sí, es decir, de las características y herramientas de las que éste dispone para gestionar los conjuntos de datos que en él están publicados.

Partiendo de esta hipótesis, esta investigación tuvo como objetivo primordial la definición de un modelo de madurez para portales de datos abiertos, valorando y definiendo para ello diferentes medidas de calidad hasta lograr establecer los criterios finales que conforman dicho modelo. Concretamente, se presenta como contribución un modelo de madurez cualitativo con varios criterios de calidad aplicables a portales de datos abiertos, considerando aspectos que no habían sido profundamente considerados en estudios o trabajos anteriores y que ameritan ser incluidos dentro de un instrumento para la medición de la madurez.

Cabe destacar que las contribuciones del trabajo de investigación desarrollado en esta tesis están siendo consideradas como parte del Programa de Gobierno Abierto que está llevando a cabo actualmente el Gobierno de Costa Rica. Al considerarse el valor agregado de contar con un modelo de madurez para portales de datos abiertos, la propuesta desarrollada en esta tesis fue utilizada como uno de los principales insumos en la creación de la norma técnica nacional INTE 27-02-02:2015 para portales de datos abiertos de Costa Rica por parte del Instituto de Normas Técnicas de Costa Rica (INTECO). Esta



norma permite evaluar portales de datos abiertos costarricenses para ubicarlos en un nivel de madurez concreto, siendo muy necesaria en el contexto nacional por lo que ha sido fundamental la incorporación del modelo generado a partir de esta investigación.

Finalmente, se debe señalar que una parte importante de este trabajo fue la aplicación piloto tanto del modelo de madurez propuesto como de la norma técnica nacional – antes mencionada – en diversos portales de datos abiertos costarricenses existentes. Como parte de las conclusiones, al final de este documento se presentan los resultados de la aplicación hecha tanto del modelo como de la norma, con el fin de dar al lector un insumo más para valorar lo presentado en esta tesis.





Índice General

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



AGRADECIMIENTOS.....	3
PREFACIO	5
CAPÍTULO 1: INTRODUCCIÓN	21
1.1 Contexto.....	24
1.2 La Web de los datos	25
1.3 Datos abiertos	27
1.4 Calidad de datos abiertos.....	30
1.5 Situación actual de datos abiertos en Costa Rica.....	33
1.6 Motivación de la investigación.....	37
1.7 Objetivos de la investigación.....	38
1.8 Contribuciones	38
CAPÍTULO 2: METODOLOGÍA	41
2.1 Investigación en Acción.....	43
CAPÍTULO 3: DATOS ABIERTOS, IMPLICACIONES Y ESTADO DE LA CUESTIÓN.....	45
3.1 Análisis DAFO sobre publicación y consumo de datos abiertos.....	48
3.2 Trabajos relacionados	49
3.2.1 Herramientas para publicación de datos abiertos.....	49
3.2.2 Modelos de calidad para datos abiertos.....	51
3.2.3 Modelos de madurez para portales de datos abiertos.....	55
CAPÍTULO 4: MODELO DE MADUREZ PARA PORTALES DE DATOS ABIERTOS.....	59
4.1 Modelo de Calidad para Portales de Datos Abiertos	61
4.1.1 Medidas de calidad propuestas	63
4.1.2 Factor de ajuste	65
4.1.3 Rangos para la evaluación final	66
4.1.4 Aplicación del modelo.....	66
4.2 Modelos de madurez	68
4.2.1 Historia del Modelo de Capacidad de Madurez (CMM)	69
4.2.2 ¿Cómo funciona el CMM?	69
4.3 Base del modelo propuesto	70
4.4 Niveles del Modelo de Madurez para Portales de Datos Abiertos	71
4.4.1 Nivel 0: Incompleto.....	71
4.4.2 Nivel 1: Básico.....	71
4.4.3 Nivel 2: Notable	71
4.4.4 Nivel 3: Destacable	71
4.5 Criterios del Modelo de Madurez para Portales de Datos Abiertos	71
4.5.1 Requisitos del criterio D/R – Disponibilidad y Reutilización	72
4.5.2 Requisitos del criterio REL – Relevancia	72

4.5.3	Requisitos del criterio REP – Reputación	73
4.5.4	Requisitos del criterio G/I – Granularidad e Integración	74
4.5.5	Requisitos del criterio V/U – Visualización y Usabilidad	74
4.6	Aplicación del Modelo de Madurez para Portales de Datos Abiertos	75
4.7	Modelo de Madurez para Portales de Datos Abiertos.....	75
CAPÍTULO 5: CASOS DE APLICACIÓN DEL MODELO DE MADUREZ PARA PORTALES DE DATOS ABIERTOS.....		77
5.1	Portal de Datos Abiertos de la Municipalidad de Pérez Zeledón	79
5.2	Portal de Datos Abiertos de la Municipalidad de Palmares	81
5.3	Portal de Datos Abiertos del MICITT	82
5.4	Conclusiones de la aplicación del modelo de madurez	85
CAPÍTULO 6: INCORPORACIÓN DEL MODELO DE MADUREZ EN LA NORMA TÉCNICA NACIONAL INTE 27-02-02:2015 PARA PORTALES DE DATOS ABIERTOS		87
6.1	Norma Técnica INTE 27-02-02:2015	89
6.2	Incorporación de los resultados de la tesis a la norma	90
6.3	Resultados de la aplicación de la norma	91
6.4	Conclusiones de la aplicación de la norma	101
CAPÍTULO 7: CONCLUSIONES Y TRABAJO FUTURO		103
7.1	Conclusiones	105
7.2	Trabajo futuro	105
CAPÍTULO 8: RESULTADOS DE LA INVESTIGACIÓN.....		107
Memoria de Investigación del DEA		109
Artículos presentados en Conferencias y Congresos		109
Artículos presentados en Revistas		110
Otras publicaciones relacionadas.....		110
CAPÍTULO 9: ANEXOS.....		111
Anexo A: Proyecto de Norma INTE 27-02-02:2015		113
Anexo B: Formulación cuantitativa del modelo de calidad presentado en (Oviedo et al 2013).....		133
Anexo C: Carta emitida por la Directora de Normalización de INTECO		137
BIBLIOGRAFÍA.....		141



Índice de Figuras

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

Figura 1: Portal Nacional de Datos Abiertos de Costa Rica	35
Figura 2: Portal de Datos Abiertos del Ministerio de Planificación de Costa Rica	35
Figura 3: Carácter cíclico de “Investigación en Acción”	43
Figura 4: Esquema de la Metodología “Investigación en Acción”	44
Figura 5: Gráfico generado en Google DSPL a partir de datos públicos	50
Figura 6: Representación gráfica del Modelo 5 Estrellas.....	53
Figura 7: Cumplimiento de la 1ª estrella	53
Figura 8: Cumplimiento de la 2ª estrella	54
Figura 9: Cumplimiento de la 3ª estrella	54
Figura 10: Cumplimiento de la 4ª estrella	54
Figura 11: Cumplimiento de la 5ª estrella	55
Figura 12: Representación del modelo de madurez de datos abiertos OD-MM	57
Figura 13: Equivalencia de los niveles del Esquema de 5 Estrellas.....	61
Figura 14: Criterios del modelo de calidad propuesto en (Oviedo et al 2013)	65
Figura 15: Niveles del Modelo de Capacidad de Madurez (CMM)	70
Figura 16: Ubicaciones geográficas en el portal de datos abiertos de la Municipalidad de Pérez Zeledón	79
Figura 17: Portal de datos abiertos de la Municipalidad de Palmares, Costa Rica	81
Figura 18: Portal de datos abiertos del MICITT	83
Figura 19: Conjuntos de datos sobre el tema TIC en el portal de datos abiertos del MICITT	83



Universitat d'Alacant
Universidad de Alicante



Índice de Tablas

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

Tabla 1: Análisis DAFO del uso de datos abiertos en una organización	49
Tabla 2: Esquema del Modelo de 5 Estrellas	52
Tabla 3: Factor de ajuste aplicable a los criterios en el modelo de calidad	66
Tabla 4: Categorías para la valoración final del nivel de calidad	66
Tabla 5: Ejemplo de aplicación del modelo en un entorno de criticidad alta	67
Tabla 6: Ejemplo de aplicación del modelo en un entorno de criticidad baja	67
Tabla 7: Esquema genérico para modelos de madurez establecido por ISO	69
Tabla 8: Evolución de los criterios del modelo de calidad al modelo de madurez	71
Tabla 9: Requisitos de Disponibilidad y Reutilización por cada nivel de madurez	72
Tabla 10: Requisitos de Relevancia por cada nivel de madurez	73
Tabla 11: Requisitos de Reputación por cada nivel de madurez	73
Tabla 12: Requisitos de Granularidad e Integración por cada nivel de madurez	74
Tabla 13: Requisitos de Visualización y Usabilidad por cada nivel de madurez	75
Tabla 14: Modelo de Madurez para Portales de Datos Abiertos	76
Tabla 15: Resultados de la aplicación del Modelo de Madurez al Portal de la Municipalidad de Pérez Zeledón	80
Tabla 16: Resultados de la aplicación del Modelo de Madurez al Portal de la Municipalidad de Palmares	82
Tabla 17: Resultados de la aplicación del Modelo de Madurez al Portal del MICITT	84



Universitat d'Alacant
Universidad de Alicante



Capítulo 1: Introducción

En este capítulo se presenta detalladamente el contexto dentro del cual se llevó a cabo la investigación que sustenta este proyecto de tesis doctoral. Además, se brinda al lector una definición general de los conceptos sobre los que se asienta el trabajo de investigación realizado, así como las relaciones, semejanzas y diferencias que existen entre dichos conceptos. Finalmente, se plantea la motivación del trabajo de investigación, así como la hipótesis y el objetivo de la misma.



Universitat d'Alacant
Universidad de Alicante

En plena era de la información, se puede afirmar que vivimos en una sociedad “datificada” (Newell & Marabelli 2015) que da una gran importancia al hecho de contar con los datos adecuados en el momento preciso con el fin de obtener el conocimiento que requerimos en nuestro día a día: para tomar las mejores decisiones, para ayudarnos en nuestro trabajo diario o en nuestro tiempo de ocio. La demanda de información por parte de las personas ha implicado que las organizaciones se esmeren cada vez más en poder contar con altos volúmenes de datos de una calidad adecuada para exponerlos públicamente a través de la Web (Heath & Bizer 2011). De hecho, el nacimiento, crecimiento y consolidación de Internet a nivel mundial ha conllevado a la necesidad de contar con instrumentos y procedimientos para la publicación de datos de calidad, con el fin de que su consumo por parte de los usuarios sea el más adecuado (van der Waal et al 2014). A lo largo de todo este tiempo en que la sociedad ha contado con la posibilidad de publicar y consultar información en la Web, se ha hecho evidente que las personas son cada vez más exigentes en cuanto a la calidad de la información que es puesta a su disposición, por lo cual se justifica la existencia de instrumentos de medición de la calidad de datos que propicien que dichas exigencias sean satisfechas. Por ello, hoy es una necesidad prácticamente de todas las organizaciones, sin importar su naturaleza o su tamaño, contar con mecanismos que aseguren una publicación de sus datos con una calidad adecuada, tanto para uso interno por parte de sus trabajadores como para uso externo por parte de sus clientes o colaboradores.

Esto resulta aún más importante en un escenario de datos abiertos (Janssen et al 2012), donde el consumo de los mismos está a disposición de todo tipo de usuarios, desde quienes solamente desean consultarlos para llevar a cabo algún análisis básico hasta quienes desean generar nuevos servicios o productos de valor añadido a partir de la reutilización de los datos accedidos, usualmente mediante el desarrollo de herramientas y aplicaciones (Janssen & Zuiderwijk 2014).

La calidad de los datos que se publican incide directamente en la calidad de los productos y servicios que se pueden elaborar a partir de ellos. De hecho, los consumidores de datos esperan encontrar siempre datos abiertos publicados con un nivel de calidad adecuado ya que los productos y servicios basados en datos que se realizan cada vez son más complejos e integrales. Un primer paso ha sido el desarrollo de portales de datos abiertos donde se establecen catálogos de datos y se ponen los medios para facilitar la publicación y consumo de conjuntos de datos (Attard et al 2015). Sin embargo, éste es sólo el primer paso, siendo necesario que los publicadores de datos puedan estimar el nivel de calidad del portal de datos abiertos y de los conjuntos de datos que contienen, a través de modelos de madurez de portales de datos abiertos que les permitan mejorar de manera continua la publicación de datos en aras de que los consumidores de datos queden satisfechos y se genere el impacto esperado.

Este capítulo introductorio consta de varias secciones. En la primera de ellas se presenta el contexto en el cual fue llevada a cabo esta investigación, describiendo las características y el conocimiento previo del autor sobre los temas que contempló en su estudio. Posteriormente se presenta el concepto de la Web de los datos, concepto relacionado con la forma en que hoy publicamos y consumimos información de manera tan versátil en Internet. Seguidamente se introducen tanto el concepto de datos abiertos como otros conceptos relacionados, para dar paso al tema de calidad de datos abiertos sobre el cual

precisamente giran las contribuciones de esta tesis doctoral. En la sección 1.5 se describe la situación actual en el tema de datos abiertos en Costa Rica, que es uno de los elementos que justamente motivan este proyecto de investigación. Se finaliza el capítulo enumerando los objetivos del proyecto y las contribuciones que éste generó a la comunidad internacional y particularmente a Costa Rica, país donde se aplican los resultados de investigación.

1.1 Contexto

El autor de esta tesis doctoral, Edgar Oviedo Blanco, es Ingeniero en Computación graduado del Instituto Tecnológico de Costa Rica, y cuenta con una maestría en Ingeniería del Software otorgada por la Universidad Politécnica de Madrid, España. El autor cuenta con amplia experiencia en el ámbito académico (por haber sido profesor universitario desde el año 1999 y contar con experiencia previa en investigación académica). Cabe destacar que en 2002 fundó Grupo Babel¹, empresa en la cual desde sus inicios se ha desempeñado como CEO, dedicada a la prestación de servicios de tecnología de información tales como desarrollo de software, implementación de portales y sistemas colaborativos, inteligencia de negocios y otros servicios afines. Esta empresa ha participado, bajo la dirección del autor de esta investigación, en varios esfuerzos individuales y colectivos relacionados al tema de implementación de buenas prácticas a favor del aseguramiento de la calidad, entre las cuales destaca el proceso de verificación de CMMi al cual fue sometida la empresa hace algunos años.

Grupo Babel cuenta con oficinas en seis países de Latinoamérica y presencia en prácticamente toda la región, siendo Costa Rica el país donde se ubica su casa matriz. Entre sus clientes se encuentran organizaciones tanto públicas como privadas, principalmente del ámbito Financiero, Agroindustrial, Comercial y Académico, destacándose las instituciones del sector gubernamental como sus principales clientes en prácticamente todos los países de la región, razón por la cual es amplio conocedor de la información que se produce y se consume en el sector público de un país. Sumado a lo anterior, Edgar Oviedo Blanco ha sido miembro de la Junta Directiva y de diversas comisiones de CAMTIC², aspectos que han contribuido significativamente a que tenga una alta comprensión sobre la realidad de las tecnologías de información en el país, incluyendo su utilización por parte de entes privados y públicos.

Este proyecto de investigación representa para el autor la tesis final para obtener el título de Doctor en Aplicaciones de la Informática en la Universidad de Alicante, España. Se ha elaborado en el seno del grupo de investigación WaKe (Web and Knowledge) del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante. Este grupo de investigación centra sus esfuerzos en el desarrollo de propuestas para mejorar la publicación y reutilización de datos abiertos, siendo el grupo que desarrolla la iniciativa de apertura de datos de la Universidad de Alicante³. Además, dado que el investigador procede del sector industrial, concretamente del sector de las tecnologías de la información

¹ <http://www.grupobabel.com>

² CAMTIC: Cámara de Empresas de Tecnología de Información y Comunicación de Costa Rica. <http://www.camtic.org>

³ <http://datos.ua.es>

en Costa Rica, y que en este país la tendencia de datos abiertos y gobierno abierto está en pleno auge, el resultado de esta investigación representa una oportunidad muy valiosa de posicionarse en el sector, como promotor y pionero en el tema de Calidad y Madurez de Portales de Abiertos.

Además, es preciso mencionar que la formación y experiencia del autor hacen factible que el escenario en el cual se desarrolló este proyecto de investigación sea eminentemente industrial, tal y como se muestra en el capítulo 2 sobre metodología, donde se describen los diversos actores que han formado parte en la investigación desarrollada. Finalmente, mediante los conocimientos del sector que tiene el autor, se espera que exista un labor de transferencia de conocimiento de I+D desde la Universidad de Alicante hacia el tejido industrial, inicialmente a partir de los resultados de este proyecto y más adelante a través de esfuerzos que eventualmente puedan formalizarse entre la Universidad de Alicante y Grupo Babel.

1.2 La Web de los datos

La dicotomía publicación-consumo de información se encuentra en el propio ADN de la Web: a finales de la década de los ochenta, Tim Berners-Lee y Robert Cailliau desarrollaron un sistema que pretendía facilitar el acceso al inmenso volumen de información que residía en los sistemas de cómputo del Consejo Europeo para la Investigación Nuclear, CERN (Berners-Lee, 1996). Este sistema se basaba en el concepto de hipertexto para compartir información y dio origen a lo que hoy se conoce como “la Web”, siendo el primer paso para que el mundo empezara a vivir a plenitud la era de la información. Poder almacenar y vincular datos de diferentes fuentes, incorporando hipertextos y multimedia en Internet, fue tan solo el inicio de un nuevo paradigma que estaría basado en el acceso ágil y simple a la información desde cualquier parte del mundo, sin importar dónde estuviese localizada físicamente su fuente.

El éxito y el crecimiento de la Web fueron más que evidentes en todos los ámbitos, destacando por supuesto la masificación de sitios que acumulaban y proveían datos a todos aquellos usuarios interesados en consultar y analizar información, ahora de una manera mucho más versátil. Proliferaron también los sitios corporativos, en los cuales empresas e instituciones de todo el mundo empezaron a proveer información acerca de sus actividades, siendo ésta – en un inicio al menos – información estática en HTML⁴ que requería de intervención experta para su modificación o actualización.

La evolución de la Web trae consigo la aparición de aplicaciones que facilitan la compartición de información de una manera mucho más dinámica, partiendo de conceptos claves como la interoperabilidad y la colaboración consolidando lo que se denominó la Web 2.0 (Vossen & Hagemann 2010). De esta manera se empezó a hacer uso masivo de la inteligencia colectiva para, de una manera sumamente interactiva, poner a disposición de los usuarios todo tipo de información, esta vez acompañada de múltiples aportes realizados por otras personas a través de herramientas como redes sociales, blogs o wikis, los cuales cambiaron radicalmente la forma clásica de ver la consulta de

⁴ **HTML**: siglas de HyperText Markup Language (lenguaje de marcado de hipertexto), que es el lenguaje de marcado predominante para la elaboración de páginas web.

información en la Web, convirtiéndola en una herramienta colaborativa más que simplemente informativa.

Los usuarios de la Web pasaron de ser solamente consumidores de información a ser también críticos y proveedores de la misma. Un ejemplo es la interactividad y otros elementos colaborativos que se agregaron en las versiones electrónicas de los periódicos, donde los lectores participan directamente con su opinión y ésta pasa a formar parte de la información que otros lectores analizan. Además, se amplió en gran medida el uso de multimedia para complementar lo que en otro momento fueron noticias compuestas exclusivamente por texto, o a lo sumo estaban acompañadas de imágenes.

La Web 2.0 fue un elemento que propició en el mundo el anhelo colectivo de poder publicar y a la vez disponer de todo tipo de información para su consumo directo o a través de servicios o productos de valor añadido (Vossen & Hagemann 2010), tales como los sistemas administradores de contenido (por ejemplo Joomla⁵) que han propiciado la creación y el crecimiento de comunidades donde se comparte información a diario, permitiendo además la incorporación de elementos avanzados de multimedia de una manera sumamente simple.

Sin embargo, la información contenida en la Web ha consistido tradicionalmente en documentos (es decir, datos no estructurados) en detrimento de contenidos estructurados, como por ejemplo podrían ser datos en un formato tabular. En (Bizer et al 2009) se expone el concepto de la "Web de los datos", la cual permite a los usuarios navegar por diferentes fuentes de datos estructurados y realizar un análisis de los mismos mediante consultas que permitan agregar o desagregar datos de la misma manera que se hace en un sistema gestor de bases de datos tradicional, con la finalidad de reutilizar los datos con el máximo potencial. Un ejemplo práctico de esto podría ser la implementación de aplicaciones móviles que permitan, gracias a una previa integración y consolidación de diversas fuentes de datos estructurados, brindar servicios de apoyo, por ejemplo:

- a) En el ámbito turístico, tal y como se ejemplifica en (Bedini et al 2014) con el caso de Islas Canarias, donde se exponen los múltiples beneficios que ha generado el uso de datos abiertos en dicho sector.
- b) En el ámbito de transportes, para conocer las mejores rutas de un origen a un destino tal y como lo hace Waze⁶, aplicación móvil que se alimenta permanentemente de la información que cada usuario le provee con sólo utilizarla.
- c) En el ámbito de la salud, para ubicar rápidamente hospitales o clínicas en casos de emergencia, tal y como lo hace "Radar CiSalud", aplicación móvil desarrollada por la Secretaría de Salud de México⁷ (disponible en las tiendas virtuales Apple Store y en Google Play).

⁵ **JOOMLA:** Plataforma que facilita el proceso de construcción de sitios web colaborativos. <http://www.joomla.org>

⁶ <http://www.waze.com>

⁷ <http://datos.gob.mx/impacto/casos-de-uso/radar-ci-salud.html>

Como puede observarse, son muchos y muy diversos los ámbitos donde la información concisa, integrada y oportuna resulta de muy alto valor para quienes la consumen.

La “Web de los datos” es la herramienta idónea para publicar los datos y permitir su consumo de una manera sencilla, siendo pieza fundamental en dotar de mecanismos para la proliferación de iniciativas de datos abiertos. Este concepto de datos abiertos no es nuevo y viene inspirado por otros conceptos similares como código abierto, acceso abierto y otros que se detallarán a continuación.

1.3 Datos abiertos

La “Web de los datos” proporciona mecanismos para publicar y consumir datos estructurados, sin embargo la característica diferenciadora de los datos abiertos, precisamente es su “libertad de acceso”. En este ámbito, e independientemente de la naturaleza de la información (estructurada o no) han existido desde hace tiempo muchas iniciativas que promueven el poder disponer libremente de la información: se trata del concepto de acceso abierto (“*open access*”). El acceso abierto tiene como objetivo “promover que cualquier usuario pueda leer, descargar, copiar, distribuir, imprimir, buscar o enlazar los textos completos de los artículos científicos, sean estos publicados en una revista, en un repositorio de acceso abierto o en una página web propia o departamental, para usarlos con cualquier otro propósito legítimo” (Wagner, 2010). Por lo tanto se centra en datos no estructurados como artículos de investigación, promoviendo el acceso al conocimiento científico sin restricciones, derivado del hecho de que éste suele estar financiado con fondos públicos y, por tanto, su acceso no debería quedar restringido por editoriales privadas bajo caras suscripciones (Abadal et al 2013).

El acceso abierto cobra vital importancia en las tareas de investigación, posibilitando una amplia difusión del conocimiento. De hecho, la filosofía de acceso abierto está potenciándose con éxito en varios países europeos con proyectos como Mediterranean Open Access Network (MedOANet⁸) cuyo objetivo es definir estrategias y políticas coordinadas sobre el acceso abierto a la información científica en los países europeos de la costa mediterránea. Por otro lado, gobiernos de todo el mundo están potenciando el acceso abierto en las contribuciones científicas, de tal manera que cualquier publicación derivada de un proyecto de investigación con financiación pública debe ser del tipo *open access*⁹.

En este sentido la relación entre acceso abierto y datos abiertos es cada vez mayor. Por ejemplo, en los últimos años varias revistas científicas (como GigaScience¹⁰) y múltiples sitios web que consolidan y ponen a disposición de la comunidad científica datos derivados de procesos de investigación (como Springer¹¹), han empezado a publicar los datos de los artículos promulgando el libre acceso de toda la información que divulgan, impulsando no sólo el acceso abierto a la publicación en sí sino a los datos de la propia investigación, todo esto con el fin de permitir a la comunidad científica la replicación o

⁸ <http://www.medoanet.eu>

⁹ <http://www.guardian.co.uk/science/2012/jul/15/free-access-british-scientific-research>

¹⁰ <http://www.springer.com/life+sciences/systems+biology+and+bioinformatics/journal/13742>

¹¹ <http://lod.springer.com>



validación de investigaciones previas en aras de potenciar un esfuerzo científico colaborativo más productivo.

Se debe mencionar que existen similitudes y diferencias entre el concepto de acceso abierto y el concepto de datos abiertos (*open data*). Como se ha comentado anteriormente, una de las diferencias estriba en la naturaleza de los datos: los datos abiertos deben ser estructurados ya que se potencia su reutilización mientras que el acceso abierto incluye información de naturaleza no estructurada. Por otro lado, la similitud entre datos abiertos y acceso abierto radica en la utilización de licencias del tipo *copyleft*, definidas en GNU.org¹² como “un método general para hacer programas u otros tipos de trabajos en modalidad libre, con la finalidad de poner a disposición de manera gratuita y sin restricciones para cierta comunidad los datos que le podrían resultar de interés”. El poder contar con una licencia es una manera de asegurar que los contenidos (ya sean datos estructurados o no) puedan ser reutilizados, potenciando que aporten valor social y económico.

Según la definición más ampliamente aceptada, dada por la OKFN¹³ (Open Knowledge Foundation), datos abiertos (*open data*) son los datos que se pueden utilizar, reutilizar y redistribuir libremente por cualquier persona, sujetos únicamente a la obligación de atribuir y compartir por igual dicha información. Para ello, los datos se publican en sitios de acceso público (conocidos como portales de datos abiertos) lo que permite, además, que estén organizados y sean fácilmente consultados.

Existe un concepto adicional llamado datos públicos (“*public data*”) que se refiere a un tipo específico de datos abiertos, y que está relacionado directamente con el sector gubernamental cuya información por su naturaleza (y en muchos casos por legislación) debe ser tratada como información de dominio público. Es decir, los datos públicos son un subconjunto de los datos abiertos que se refiere específicamente a aquella información que proviene de fuentes gestionadas por organizaciones que forman parte de la administración pública y que, tal y como se menciona en (Jetzek et al, 2012), se habían acumulado durante siglos en repositorios protegidos.

La información que comúnmente se puede observar en los sitios que muestran datos públicos (portales de datos abiertos) por lo general está relacionada con el empleo de los recursos públicos, la seguridad nacional, la educación, la salud, estadísticas socio-demográficas y otros temas que, en conjunto, dan paso al concepto de “gobierno abierto” (*open government*). Estos sitios ponen a disposición de sus visitantes herramientas tales como foros, blogs o chats, los cuales propician mediante la participación ciudadana un amplio intercambio de información y generan en la sociedad un sentimiento de transparencia por parte de los entes gubernamentales.

Particularmente el concepto de gobierno abierto ha tenido una fuerte acogida por parte de diversos países de todas partes del mundo, gracias a que mediante la publicación de la información – ya sea de

¹² <http://www.gnu.org/licenses/copyleft.es.html>

¹³ <http://okfn.org/opendata>

alcancel nacional o regional – potencian la transparencia y la rendición de cuentas, dos conceptos que, por diversas razones y realidades, cada vez son más demandados por los ciudadanos.

A pesar de la falta de legislación en algunos gobiernos para facilitar el libre acceso a algunos segmentos de información, los datos públicos han ido tomando cada vez más fuerza alrededor del mundo. De hecho, hoy día existen múltiples países que cuentan con sus propios sitios web dedicados a facilitar el acceso a datos públicos bajo una legislación específica. Como ejemplos se pueden citar:

- <http://data.gov>, portal de datos abiertos del Gobierno de los Estados Unidos.
- <http://data.gov.uk>, portal de datos abiertos del Reino Unido, en el cual el propio Primer Ministro, David Cameron, invita por medio de un vídeo a los visitantes a hacer uso ilimitado y a explotar todo el contenido del sitio.
- <http://datos.gob.es>, portal de datos abiertos del Gobierno Español, que está amparado por el Real Decreto 1495/2011, de 24 de octubre, del cual deriva la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público para el ámbito del sector público estatal¹⁴, modificada recientemente mediante la Ley 18/2015, de 9 de julio¹⁵.

En América Latina también existe un progreso considerable en cuanto a la publicación de portales gubernamentales de datos abiertos, siendo un evidente ejemplo de ello el gran avance mostrado por Chile, país que desde hace varios años cuenta con un portal de Gobierno Abierto¹⁶ y ofrece a la ciudadanía un kit digital¹⁷ que incluye tanto la norma técnica (aprobada en 2013) para la publicación de datos abiertos como una guía rápida de publicación de datos abiertos. En esta región del mundo también existen esfuerzos formales por implementar ampliamente el concepto de Gobierno Abierto en países como Argentina (www.gobiernoabierto.gob.ar), México (www.gobabiertomx.org), El Salvador (www.gobiernoabierto.gob.sv) y muchos otros, entre ellos Costa Rica cuya situación actual con respecto a datos abiertos se expone con detalle en la sección 1.5.

Considerando la tendencia de crecimiento en la publicación de datos abiertos alrededor del mundo, es evidente la necesidad de contar con mecanismos para garantizar la calidad de estos datos, partiendo de que la calidad de los productos o servicios de valor añadido realizada a través del consumo de datos abiertos (desde aplicaciones hasta análisis de los datos) depende directamente de la calidad de los datos abiertos publicados en los portales de datos abiertos. Esto lo han percibido en muchos gobiernos que ya han emprendido diferentes esfuerzos por estandarizar y regular la calidad de los portales de datos abiertos que se publiquen en sus países.

¹⁴ <http://www.boe.es/boe/dias/2011/11/08/pdfs/BOE-A-2011-17560.pdf>

¹⁵ <https://www.boe.es/boe/dias/2015/07/10/pdfs/BOE-A-2015-7731.pdf>

¹⁶ <http://www.gobiernoabierto.cl>

¹⁷ <http://kitdigital.gob.cl/recursos-de-desarrollo>

1.4 Calidad de datos abiertos

Un ejemplo paradigmático de proyecto de apertura de datos públicos que considera aspectos de calidad es el proyecto APORTA¹⁸, impulsado por el Ministerio de Industria, Energía y Turismo y el Ministerio de Hacienda y Gobernaciones Públicas, ambos del Gobierno Español, debido a que en este proyecto se llevó a cabo un trabajo de recopilación de conclusiones, síntesis y constitución de un decálogo, como elemento fundamental a tener en cuenta para potenciar la filosofía de datos abiertos en España¹⁹. Este decálogo, relacionado directamente con el concepto de datos públicos, justamente busca garantizar que los datos publicados tengan un nivel de calidad que satisfaga las expectativas de los consumidores de datos. Un breve resumen de este decálogo se detalla a continuación:

0. Armonización entre administraciones.

Todos los puntos del decálogo se basan en la premisa de que debe existir una armonización entre todas las administraciones públicas. Todas las iniciativas de datos abiertos deben compartir los mismos principios y definiciones que se listan en el decálogo. Este punto 0 es básico para la interoperabilidad y aprovechamiento eficiente de las sinergias llevadas a cabo por todos los actores datos abiertos – RISP (Reutilización de Información en el Sector Público).

1. Publicar datos en formatos abiertos y estándares.

Cualquier iniciativa de datos abiertos debería publicar sus conjuntos de datos en formatos abiertos (no propietarios) y que sean adecuados para permitir la reutilización de los mismos por parte del colectivo destinatario.

2. Usar esquemas y vocabularios consensuados.

Además de los formatos abiertos y estándares, la estructura de los datos debería seguir un convenio o unos esquemas definidos. Si se crean vocabularios o esquemas de representación de la información específicos, éstos se deberían exponer públicamente para que se pueda interpretar correctamente la información al reutilizarla.

3. Inventario en un catálogo de datos estructurado.

Cualquier iniciativa de datos abiertos debe tener un punto de consulta donde se incluya un inventario con información descriptiva y técnica sobre los conjuntos de datos que se exponen. Los metadatos que informan sobre cada conjunto de datos deberían seguir una estructura común y estándar.

4. Datos accesibles desde direcciones web persistentes y amigables.

Tanto las fichas de los conjuntos de datos, como la distribución de la propia información (volcado en un archivo, API de consulta, RSS, etc.) deberían de estar accesibles desde URLs que persistan en el tiempo y así evitar que se pierdan las referencias en el futuro. Además, deben seguir una estructura homogénea y bien definida, con información legible con el fin de facilitar su reutilización mediante el conocimiento del contenido referido por dichas URLs.

¹⁸ <http://datos.gob.es>

¹⁹ <http://datos.gob.es/content/decalogo-del-reutilizador-de-datos-abiertos-del-sector-publico>



5. Exponer un mínimo conjunto de datos relativos al nivel de competencias del organismo y su estrategia de exposición de datos.

Cada administración que impulse una iniciativa de datos abiertos debería crear una hoja de ruta donde especifique la estrategia de exposición de los conjuntos de datos y sus prioridades. Inicialmente, debería publicar los conjuntos de mayor interés según las competencias del propio organismo.

6. Compromiso de servicio, actualización y calidad del dato, manteniendo un canal eficiente de comunicación entre reutilizador y administraciones públicas.

La administración debe mantener un mínimo de calidad y servicio en su iniciativa de datos abiertos, manteniendo lo expuesto en la estrategia de publicación y comprometiéndose con su colectivo reutilizador. Debe establecer un canal eficiente de comunicación que permita la interacción bidireccional entre organismo público y reutilizadores.

7. Monitorizar y evaluar el uso y servicio mediante métricas.

La administración debe crear métricas y evaluar sus indicadores de uso y servicio de la iniciativa de datos abiertos. De esta forma puede monitorizar el funcionamiento y uso, y así analizar si se está cumpliendo el compromiso con la comunidad de reutilizadores y cuáles son las potenciales carencias del sistema o de la estrategia.

8. Datos bajo condiciones de uso no restrictivas y comunes.

Las condiciones de uso deberían ser lo menos restrictivas posible y permitir la reutilización libre, incluso para fines comerciales. Se recomienda la creación y uso de licencias tipo autodocumentadas (por ejemplo ODbL²⁰) y que sean comunes entre distintas administraciones.

9. Evangelizar y educar en el uso de datos.

Es necesario educar en el uso de los datos, tanto a los colectivos de reutilización específicos (sector TIC, periodismo, investigación, etc.) como a la sociedad en general y así fomentar el conocimiento y la inquietud por procesar información de una forma autónoma.

10. Recopilar aplicaciones, herramientas y manuales para motivar y facilitar la reutilización.

Cualquier iniciativa de datos abiertos debería recopilar ejemplos de uso y herramientas que faciliten y motiven la reutilización de los datos que se publican.

Además de este decálogo, en relación a la calidad de los datos, se debe señalar que, en muchos casos, los datos abiertos (incluidos los datos públicos) provienen de una fuente de datos que se diseñó con un objetivo inicial diferente a la elaboración de productos y servicios de valor añadido a través de su reutilización. De hecho, estas fuentes de datos usualmente tendrían un objetivo englobado dentro de la operativa diaria de la organización, cubriendo sus requisitos funcionales. Por ello, debe considerarse que cada uno de los diferentes conjuntos de datos debe prepararse antes de publicarlo, lo cual requiere un proceso que incluye integración y homogenización de los datos a publicar. También debe

²⁰ <http://opendatacommons.org/licenses/odbl>

considerarse que, tradicionalmente, buena parte de la información había estado bajo control de organizaciones que habían restringido su acceso, muchas veces por medio de licencias de uso o derechos de autor (*copyright*). Por todo lo anterior, el grado de eficacia con que se lleve a cabo la preparación de los datos a publicar en un portal de datos abiertos tendrá gran inherencia en el grado de calidad que perciban las personas que los consuman, ya sea a través de los portales de datos abiertos o a través de herramientas habilitadas para su procesamiento fuera del portal.

Estas iniciativas son interesantes, si bien insuficientes, ya que debemos estudiar de una manera más completa la calidad de los datos abiertos. Para ello, se debe estudiar el concepto de calidad de datos en sí: calidad de datos es un concepto que tradicionalmente se ha utilizado para referirse a un conjunto de técnicas, que permiten garantizar la fidelidad y consistencia de los datos que son considerados como entradas de un proceso de análisis de información (Wang y Strong 1996). Sin embargo, es necesario considerar otra serie de conceptos relevantes que indudablemente están vinculados con el nivel de calidad de un conjunto de datos que se publican. Por ello, para hablar de calidad de datos resulta indispensable considerar el concepto de adecuación al uso (o *"fitness for use"* en inglés), que se refiere a que los datos deben cumplir con varios requisitos para poder emplearse de manera útil, cumpliendo un objetivo específico en un determinado contexto (Wang y Strong 1996).

Con la aplicación de las técnicas de calidad de datos que tengan en cuenta la adecuación al uso se garantiza que la información posea la calidad necesaria, no gastando esfuerzos innecesarios en asegurar una calidad de datos que no sea finalmente percibida por los usuarios.

Por ejemplo, en la implementación de una solución de inteligencia de negocios para el sector sanitario es fundamental que los datos sean muy precisos cuando se integren de diferentes fuentes, sobre todo si se trata del historial clínico de un paciente. Por supuesto, también es muy importante la seguridad con que se manejen esos datos. Sin embargo, en este caso no tiene tanta relevancia la reputación como criterio para definir la calidad de los datos, pues se parte de que las fuentes de donde proviene la información son 100% confiables. Por el contrario, para casos donde las fuentes de datos sean menos controladas (como el caso de los portales de datos abiertos) será necesario adoptar criterios distintos a los del ejemplo anterior, entre los cuales podría estar la visualización de los datos publicados. En (Graves y Hendler 2013) se recalca la importancia de la facilidad de visualización en todo portal de datos abiertos, justificando esto en que hay una importante porción de la población que no se beneficia de los datos abiertos al no poder realizar las operaciones esenciales para recolectar, procesar, combinar e interpretar los datos debido a la falta de experticia y de conocimiento técnico, situación que sugieren corregir proveyendo mecanismos simples para entender y comunicar grandes volúmenes de datos.

Cuando se habla acerca de adecuación al uso en datos abiertos, significa que los datos abiertos publicados deben garantizar que se puedan consumir para poder desarrollar productos y servicios de valor añadido a partir de ellos.

Calidad de datos abiertos se refiere al grado en que los datos publicados cumplen con criterios específicos, previamente establecidos y regulados, que giran alrededor de diferentes aspectos como el

origen de los datos, la segmentación e integración de los mismos, el grado de especificidad o generalidad con que se manejen, la visualización que se haga de los mismos, y otros. Por ello se puede decir que un dato abierto es un dato que cumple con ciertos criterios de calidad y que, por tanto, puede ser más o menos abierto en relación al modelo de calidad con que sea evaluado. Además, según el concepto de adecuación al uso anteriormente descrito, deben existir criterios adicionales que estén en función del uso real que se dé a los datos, evaluando así la calidad de los datos a lo largo de los procesos en los que se utilizan y proporcionando medidas subjetivas (Batini et al 2009).

Por ello, y debido a que las fuentes y los contextos en los que se publican datos abiertos son tan diversos, se vuelve transcendental definir un modelo que permita establecer si estos cumplen con los requerimientos mínimos para ser considerados como datos abiertos de calidad (Mazón et al. 2012). Sin embargo, es necesario determinar primero qué significa que un dato sea abierto y de calidad. Esta investigación está dirigida precisamente hacia ese objetivo: crear un modelo de madurez que permita verificar el grado de cumplimiento de ciertos criterios de calidad tanto en los datos publicados en un portal de datos abiertos como en las herramientas provistas para su consumo y análisis dentro de dicho portal.

Más adelante en este documento se presenta una definición más amplia y concisa sobre este concepto, así como las medidas que regulan a cada uno de los aspectos antes mencionados.

1.5 Situación actual de datos abiertos en Costa Rica

En los últimos años, gracias al surgimiento y fortalecimiento de iniciativas a favor de una mayor apertura y disponibilidad de la información que se publica en Internet, respaldadas además por nueva legislación emitida al respecto en muchos países, ha crecido el afán (y con ello la tendencia) de ir eliminando las restricciones y los mecanismos de control para el acceso a los datos estructurados, y también su uso, reutilización e interpretación. Siguiendo esta tendencia mundial hacia la apertura de datos, en enero del 2012 la República de Costa Rica se incorporó a la iniciativa multilateral conocida como Alianza para el Gobierno Abierto (*Open Government Partnership* u OGP por sus siglas en inglés).

Gobierno Abierto²¹, cuyos objetivos radican en la divulgación masiva de información pública basada en los principios de transparencia, participación ciudadana y rendición de cuentas, es un nuevo paradigma de gestión pública que busca mejorar los niveles de transparencia y acceso a la información, facilitar la participación ciudadana en la formulación, implementación y evaluación de política pública, y propiciar la generación de espacios de trabajo colaborativo. Como elementos diferenciadores de las estrategias para lograrlo están el involucramiento de la sociedad civil y la ciudadanía, y el uso de herramientas novedosas basadas en las Tecnologías de la Información y Comunicación (TIC).

Como parte de la OGP, Costa Rica presentó en abril del 2013 su primer plan de acción en la materia, documento que fue evaluado por el Mecanismo de Reportes Independientes (IRM) en febrero del 2015. Siguiendo el cronograma establecido por la OGP, Costa Rica debía elaborar un segundo plan de acción

²¹ Programa Gobierno Abierto de la República de Costa Rica: <http://presidencia.go.cr/gobiernoabierto-2/>

para el período 2015-2017 que estableciera sus compromisos para aumentar la transparencia, fomentar la participación ciudadana y promover espacios de colaboración cívica. Esta labor, que incluyó la participación de diversas organizaciones civiles²², inició formalmente el 20 de mayo del 2015 con el evento que se denominó “Lanzamiento del proceso de elaboración del Plan de Acción de Gobierno Abierto 2015-2017 Costa Rica”, en el cual participó el autor de esta tesis doctoral en representación del Instituto de Normas Técnicas de Costa Rica (INTECO), que desde inicios del mismo año había iniciado el desarrollo de una norma técnica para verificar la calidad y la madurez de los portales de datos abiertos (de esta norma se habla ampliamente en los capítulos posteriores de este documento). Finalmente, la primera versión oficial del Plan de Acción de Gobierno Abierto se obtuvo en octubre de 2015 y fue oficialmente presentado en diciembre de 2015 por parte del Gobierno de la República de Costa Rica.

Además, dentro del mismo contexto, el Gobierno de Costa Rica publicó el Decreto Ejecutivo N° 38994-MP-PLAN-MICITT mediante el cual se crea la Comisión Nacional por el Gobierno Abierto, con el objetivo de: “Fomentar los principios del Gobierno Abierto en la Administración Pública de Costa Rica, principios que se manifiestan en: mejorar los niveles de transparencia, garantizar el acceso democrático a la información pública, promover y facilitar la participación ciudadana e impulsar la generación de espacios de trabajo colaborativo interinstitucional y ciudadano; mediante la innovación y aprovechando al máximo las facilidades que brindan las Tecnologías de la Información y Comunicación (TIC)”. Esta comisión promueve políticas, lineamientos, estrategias y una metodología de evaluación en materia de Gobierno Abierto, y coordina las acciones necesarias para cumplir los principios de transparencia, rendición de cuentas y participación ciudadana. Está conformada por varios ministros y viceministros, dos representantes de la sociedad civil, un representante del Consejo Nacional de Rectores (CONARE) y un representante del sector empresarial.

Existen en Costa Rica diversos portales de datos abiertos, principalmente de ministerios, instituciones autónomas y municipios, que están disponibles para que los ciudadanos se enteren con absoluta transparencia de aspectos socialmente sensibles tales como la inversión de fondos públicos, los proyectos en desarrollo y todo tipo de información propia del entorno al que se refiere la organización. Algunos ejemplos de portales de datos abiertos existentes en Costa Rica son:

- **Portal Nacional de Datos Abiertos**

<http://datosabiertos.gob.go.cr>

Es el portal oficial de datos abiertos del Gobierno de Costa Rica. Integra y permite acceder información de más de 20 instituciones públicas participantes en la iniciativa de Gobierno Abierto.

²² Entre las organizaciones civiles que han tenido participación activa en el desarrollo del Plan de Acción de Gobierno Abierto de la República de Costa Rica están: Abriendo Datos Costa Rica (<http://abriendodatoscostarica.org>), Asociación Centro Ciudadano de Estudios para una Sociedad Abierta (ACCESA <http://accesa.org>) y Estudio Manatí (<http://www.estudiomanati.com>).



Figura 1: Portal Nacional de Datos Abiertos de Costa Rica

- **Portal de Datos Abiertos del Ministerio de Planificación**

<http://datosabiertos.mideplan.go.cr>

Este portal de datos abiertos muestra diversos tipos de indicadores relacionados con los tópicos que supervisa este ministerio, tales como asuntos sociales, asuntos internacionales y otros.



Figura 2: Portal de Datos Abiertos del Ministerio de Planificación de Costa Rica



- **Portal de Datos Abiertos del Ministerio de Ciencia y Tecnología**

<http://indicadores.micit.go.cr>

Este portal de datos abiertos, considerado uno de los mejores del país, muestra de manera muy amplia y diversa los indicadores relacionados con los temas de interés en el ámbito de la ciencia, la tecnología y las telecomunicaciones. En el capítulo 5 se explica con mayor amplitud este portal.

- **Portal de Datos Abiertos de la Municipalidad de Pérez Zeledón**

<http://opendata.mpz.go.cr>

Este portal de datos abiertos fue desarrollado en la etapa inicial de esta investigación con el objetivo de aplicar el modelo de calidad que fue propuesto en (Oviedo et al 2013). En el capítulo 5 se explica con mayor amplitud este portal.

- **Portal de Datos Abiertos de la Municipalidad de Palmares**

<http://datos.munipalmares.go.cr>

Este portal de datos abiertos es considerado uno de los más completos entre todos los municipios de Costa Rica, debido a la gran diversidad y cantidad de indicadores que muestra. En (Oviedo et al 2015) se muestran los resultados de la primera aplicación del modelo de madurez que se propone en este documento. En el capítulo 5 se explica con mayor amplitud este portal.

Tanta es la importancia que se le está dando a los datos abiertos en Costa Rica que el Instituto de Normas Técnicas de Costa Rica²³ (INTECO) incursionó en el desarrollo de una norma técnica sobre esta temática. INTECO es una asociación privada, sin fines de lucro, con personería jurídica y patrimonio propio. Fue creada en 1987 y reconocida en el año 1995, por decreto ejecutivo, como el Ente Nacional de Normalización, hecho que se consolida con la emisión de Ley del Sistema Nacional para la Calidad, N° 8279, publicada el 21 de mayo del 2002.

Los objetivos de INTECO son:

- Liderar la elaboración de normas costarricenses, asegurando que sean convenientes para el desarrollo socio-económico del país, que su preparación es conforme con las buenas prácticas de normalización internacionalmente aceptadas y que promuevan el mejoramiento de la calidad de los procesos, productos y servicios diseñados, fabricados, transformados, utilizados o vendidos en el país, sean ellos nacionales o importados.
- Fomentar y desarrollar actividades de certificación articuladas con los procedimientos de evaluación de la conformidad.
- Colaborar con las entidades del Sector Público con el fin de alcanzar la mayor implantación y utilización de las actividades de normalización y certificación.
- Promover la participación de Costa Rica en las organizaciones internacionales y regionales de normalización y certificación, desempeñando en las mismas la representación, en los términos que corresponda, y asumiendo los derechos y obligaciones que lleva aparejados dicha representación.

²³ INTECO – Instituto de Normas Técnicas de Costa Rica. <http://inteco.or.cr/esp/acerca-de-inteco>



- Llevar a cabo todas aquellas actividades que – relacionadas con la normalización y certificación – contribuyan a mejorar su conocimiento, utilización y desarrollo en la sociedad, así como a favorecer los intercambios comerciales, la cooperación internacional, y permitan la generación de resultados positivos que garanticen un desarrollo sostenible de la Asociación.

A pesar de contar con varias y muy diversas normas técnicas relacionadas con temas informáticos y de desarrollo de software, hasta el año 2014 INTECO no contaba aún con una norma que tuviera inherencia directa en la calidad de los portales de datos abiertos, algo que la industria local estaba requiriendo para garantizar que el país contara con portales de datos abiertos que mostraran niveles homogéneos de calidad. Dado lo anterior y dada la relevancia que estaban empezando a tener en Costa Rica los temas de gobierno abierto y de datos abiertos, a finales de ese mismo año INTECO analizó la opción de crear una norma técnica nacional para portales de datos abiertos, a través de un subcomité creado meses atrás para atender temas relacionados con la temática de Ingeniería de Software, bajo la supervisión de la ingeniera Francesca Rappaccioli (funcionaria de INTECO).

En el desarrollo de esta norma se lleva a cabo un esfuerzo conjunto entre representantes de varias instituciones del Gobierno, empresas y organizaciones del Sector Privado, instituciones del Sector Académico y otros gremios relacionados. A lo largo de todo el proceso de la elaboración de la norma, el subcomité que se hizo cargo fue presidido por el autor de este trabajo de investigación, ingeniero Edgar Oviedo Blanco, quien durante el año 2015 incorpora en el desarrollo de la norma los distintos resultados de investigación obtenidos de este trabajo de tesis doctoral, resultando de especial relevancia el Modelo de Madurez para Portales de Datos Abiertos.

1.6 Motivación de la investigación

La necesidad latente de las personas de contar con información completa, actualizada y confiable de manera ágil e inmediata, así como la legislación vigente que impone a los gobiernos (centrales y locales) la obligación de divulgar abiertamente toda la información de naturaleza pública, han conllevado a un fuerte aumento en la implementación de portales de datos abiertos por parte de todo tipo de organizaciones alrededor del mundo. Ante esto surge la necesidad de garantizar un nivel de calidad mínimo y adecuado tanto en los datos que se publiquen como en las herramientas informáticas (portales) que se diseñen para permitir su acceso e interpretación por parte de los usuarios.

El uso de datos abiertos propicia la inclusión de muy diversas fuentes de información, desde fuentes muy formales y estructuradas como las que puede proveer una organización (sea pública o privada) hasta fuentes más informales estructuradas de alguna forma básica (por ejemplo, una tabla en un documento realizado con un procesador de textos). Ante ello, surge una serie de incógnitas relacionadas con la calidad de los datos abiertos, tales como: ¿Qué criterios de calidad aparecen, según la adecuación al uso en los datos abiertos? ¿Son estos criterios de calidad específicos de los datos abiertos? ¿Cómo se considera la calidad de los portales donde están integrados los datos abiertos? ¿Cómo se puede medir la calidad de los datos que se publican como datos abiertos?

La exigencia por parte de los consumidores de datos abiertos de una calidad adecuada a la reutilización de los datos, hace que resulte necesario definir un instrumento estándar de medición que permita valorar integralmente los portales de datos abiertos, ubicándolos en una escala de madurez a partir del cumplimiento de los criterios de calidad establecidos y permitiendo a las organizaciones que los publican conocer cuáles son los criterios que permitirán que su portal sea satisfactorio, y que se englobe dentro de un proceso de mejora continua para que, eventualmente, se escale a los siguientes niveles de madurez.

1.7 Objetivos de la investigación

Este proyecto de investigación tiene como objetivo principal crear un modelo que permita evaluar portales de datos abiertos con el fin de ubicarlos en un nivel de madurez acorde al cumplimiento de ciertos criterios de calidad establecidos. Estos criterios de calidad se han definido según el nivel de adecuación de los datos a la reutilización, estableciendo una propuesta concreta para la evaluación de la calidad de portales de datos abiertos. Para lograr el cumplimiento de este objetivo principal se han definido los siguientes objetivos secundarios:

- Analizar los trabajos previos realizados sobre los conceptos que involucra esta investigación, tales como datos abiertos, calidad de datos, modelos de calidad o modelos de madurez.
- Definir los criterios de calidad que permitan realizar un análisis exhaustivo de la calidad de un portal de datos abiertos.
- Estudiar e incorporar en el modelo propuesto el concepto de *“fitness for use”* (adecuación al uso), aplicado a los criterios de calidad que se consideren al medir la calidad de un portal de datos abiertos.
- Organizar los criterios de calidad seleccionados de manera estructurada y evaluable dentro de un modelo que permita, a partir de estos, determinar la madurez de un portal de datos abiertos.
- Aplicar el modelo de madurez propuesto a diferentes portales de datos abiertos, de modo que sea posible analizar sus resultados y eventualmente mejorarlo en próximas versiones.
- Incorporar el modelo de madurez desarrollado a la norma técnica de calidad de Costa Rica desarrollada por INTECO y aplicar dicha norma a un portal de datos abiertos de alta relevancia para ejemplificar el uso del modelo de madurez en la medición de la calidad de los portales de datos abiertos en Costa Rica. Esto además servirá para valorar sus resultados y establecer mejoras al modelo y a la norma.

1.8 Contribuciones

Considerando que la calidad de un portal de datos abiertos debe contemplar aspectos tanto de la calidad de datos como de la calidad del portal en sí, es decir, de las características y herramientas de las que éste dispone para gestionar los datos que en él están publicados, la propuesta de solución se basa en definir un modelo de madurez para portales de datos abiertos que incluya criterios de calidad de ambos tipos: calidad de datos abiertos y calidad del portal de datos abiertos.



En cuanto a la calidad de los datos, ha sido ampliamente planteado por diversos autores (Pipino et al 2002, Redman 1998) que ésta determina en un alto porcentaje la calidad de los resultados que se obtengan al analizar dichos datos. Es necesario tener muy presente que la calidad de datos abiertos difiere de la calidad de datos tradicionales, ya que en el primer escenario no se está en un entorno controlado como el que se tiene al manipular datos privados de una organización. Por ejemplo, entre los criterios que deben considerarse en el análisis de la calidad de datos abiertos está la reputación de sus fuentes, criterio que no sería tan relevante en un análisis tradicional de datos de una empresa, generados a lo interno de ella.

En cuanto a la calidad del portal donde se publican e interpretan los datos abiertos, debe considerarse la necesidad de evaluar con detalle la estructura del portal junto con todas sus características y funcionalidades, a fin de determinar si representa una herramienta apropiada para un usuario que desea acceder y analizar datos abiertos publicados en él.

Los resultados finales que se esperan producto de esta investigación son:

- a) Un modelo para medir la madurez de los portales de datos abiertos basado en criterios de calidad de datos abiertos.
- b) La incorporación del modelo de madurez propuesto en esta investigación dentro de la norma técnica desarrollada por INTECO en Costa Rica, de manera que la evaluación de madurez de los portales que se someten a la norma se basen en dicho modelo.



Universitat d'Alacant
Universidad de Alicante



Capítulo 2: Metodología

En este capítulo se exponen las características y los detalles de la metodología que se adoptó para poder llevar a cabo esta investigación, la cual se denomina “Investigación en Acción”.

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

En todo trabajo de investigación es preciso seguir algunos lineamientos metodológicos que permitan asegurar cierto nivel de satisfacción en el resultado final. Entre los diversos métodos de investigación cualitativa existentes, uno de los más usados en sistemas de información e ingeniería del software es Investigación en Acción (*Action Research* en inglés). A continuación se describen los aspectos más importantes de este método presentado en (Genero et al, 2014).

2.1 Investigación en Acción

Debido a la naturaleza de la investigación llevada a cabo (investigación aplicada), la aproximación metodológica que se ha seguido es la de "Investigación en Acción" (McTaggart 1991, French & Bell, 1996; Wadsworth, 1998; Avinson et al, 1999; Seaman, 1999; Padak & Padak, 1994). Este proceso de investigación no es un proceso lineal, sino que va avanzando mediante la ejecución de ciclos, en cada uno de los cuales se ponen en marcha nuevas ideas que son puestas en práctica, comprobadas y refinadas en el siguiente ciclo, como se muestra en la Figura 2. Los pasos que se siguen en las investigaciones que utilizan este método son:

- Planificación: donde se identifican las cuestiones que guiarán la investigación.
- Acción: donde se recoge la información relacionada con tales cuestiones.
- Observación: con el fin de analizar la información recogida.
- Reflexión: que consiste en compartir los resultados con el resto de interesados, de tal manera que se invite al planteamiento de nuevas cuestiones relevantes a proporcionar conocimientos nuevos y obtener soluciones refinadas.



Figura 3: Carácter cíclico de "Investigación en Acción"

Una característica importante de esta metodología es que debe contar con una serie de actores con el fin de facilitar su aplicación (ver Figura 3). En esta investigación, los actores han sido los siguientes:

- Investigadores: el autor de esta memoria de investigación y el resto de integrantes del grupo de investigación WaKe (Web and Knowledge) del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante (<http://wake.dlsi.ua.es>).
- Objeto investigado: modelo de madurez para portales de datos abiertos.
- Grupo crítico de referencia:
 - JUNAR (www.junar.com), empresa privada que provee una plataforma para el desarrollo y la gestión de portales de datos abiertos. Actualmente es una plataforma muy extendida en América Latina, siendo la plataforma más empleada en Costa Rica para este fin, por lo cual será altamente evaluada con cada uno de sus portales a los que les sea aplicado el modelo de madurez derivado de este trabajo de investigación.
 - Grupo BABEL (www.grupobabel.com), empresa privada dedicada a servicios de desarrollo de software, inteligencia de negocios e implementación de portales de contenido en tecnologías Microsoft. Dado el crecimiento en la publicación de datos abiertos en Costa Rica, se prevé que muchas organizaciones opte por utilizar tecnologías de Microsoft para implementar sus portales de datos abiertos.
- Partes implicadas o beneficiarios:
 - INTECO - Instituto de Normas Técnicas de Costa Rica (www.inteco.or.cr).
 - Municipalidad de Pérez Zeledón (www.mpz.go.cr).
 - Municipalidad de Palmares (www.munipalmares.go.cr).
 - Ministerio de Ciencia, Tecnología y Telecomunicaciones (www.micitt.go.cr).



Figura 4: Esquema de la Metodología "Investigación en Acción"



Capítulo 3: Datos abiertos, implicaciones y estado de la cuestión

En este capítulo se abarca de manera detallada el concepto de datos abiertos, incluyendo varias definiciones afines recopiladas de diversos trabajos previos, que permiten tener un más amplio panorama de las implicaciones del término alrededor del cual gira este trabajo de investigación. Además, se incluye un análisis DAFO sobre la publicación y consumo de datos abiertos.

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

Dado el crecimiento exponencial de la información disponible en las miles de fuentes de datos alrededor del mundo, han surgido múltiples esfuerzos por integrar y clasificar dichos datos, posibilitando su publicación y posterior análisis ya sea con fines privados o por un beneficio social. Como se comentó brevemente en la introducción, dentro de este amplio contexto tomó mucha fuerza un movimiento impulsado por Tim Berners-Lee que consistió en un llamamiento a compartir datos libremente mediante el uso de la Web para el beneficio de toda la comunidad²⁴, así como en algunos resultados interesantes obtenidos un tiempo después de dicho llamamiento²⁵. Se debe entender como datos abiertos aquellos datos que se publican en sitios web de acceso público con el fin de ser reutilizados y redistribuidos por toda persona u organización que lo desee, sin tener ningún tipo de restricciones para ello. Los datos abiertos suelen estar disponibles a través de portales web que permiten su análisis con el fin de propiciar una toma de decisiones informada.

En los últimos años el tema de datos abiertos ha venido adquiriendo mayor interés por parte de las autoridades y de la población en general, en todo el mundo. Este innovador pero a la vez sencillo concepto se ha popularizado gracias a la necesidad y el deseo que tienen las personas de poder acceder a información fidedigna, completa y de manera oportuna relacionada con temas de interés general, tales como el uso de presupuestos públicos, proyectos de desarrollo económico y social de los pueblos, información demográfica, entre otros.

A pesar que ha mostrado un mayor apogeo en varios países de Europa (como el Reino Unido) y en los Estados Unidos, en Latinoamérica aunque el crecimiento en el uso de datos abiertos ha sido apenas moderado, los gobiernos centrales y las instituciones estatales han mostrado gran interés en implementar todo tipo de herramientas informáticas que permitan impulsar, principalmente, dos conceptos que resultan de vital importancia para los gobernantes: transparencia y rendición de cuentas. De hecho, en (Abella et al, 2014) se refuerza esta idea al indicar que “las políticas para el aumento de la transparencia, la implantación del gobierno abierto o de las ciudades inteligentes tienen en el open data uno de sus pilares fundamentales”.

Ese creciente interés por contar con portales web que muestren y permitan reutilizar todo tipo de información que por su naturaleza es pública (conocidos como “portales de datos abiertos”) ha permitido inclusive el surgimiento de iniciativas formales al respecto en algunos países de la región. Un ejemplo de ello es Chile, país donde fue desarrollada y está rigiendo desde febrero de 2013 una norma técnica para la publicación de datos abiertos²⁶, actualmente en su versión 2.1. Esto mismo, e incluso de forma más amplia, ha sido implementado en varios países de Europa, siendo España ejemplo de un país

²⁴ http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html

²⁵ http://www.ted.com/talks/tim_berniers_lee_the_year_open_data_went_worldwide.html

²⁶ http://instituciones.gobiernoabierto.cl/NormaTecnicaPublicacionDatosChile_v2-1.pdf

donde el concepto de datos abiertos se incluyó dentro de un concepto mayor llamado “ciudades inteligentes”, para lo cual incluso se publicó el Plan Nacional de Ciudades Inteligentes²⁷.

La implementación de portales de datos abiertos propicia la inclusión de datos provenientes de muy diversas fuentes de información, desde fuentes muy formales y estructuradas (como las que podría proveer una organización dedicada a investigaciones y censos), hasta fuentes más informales que podrían brindar información organizada de manera muy básica. Por ello, es necesario estudiar a fondo las diferentes características de un portal de datos abiertos para establecer aquellos criterios que permitan justamente evaluar su calidad.

3.1 Análisis DAFO sobre publicación y consumo de datos abiertos

Con la finalidad de comprender mejor las ventajas y desventajas que representa para una organización promover la publicación de datos abiertos y como primer paso en la búsqueda de criterios de calidad, se realizó un análisis del tipo DAFO (Debilidades / Amenazas / Fortalezas / Oportunidades). Las fortalezas y debilidades contemplan factores internos que pueden ser controlables por la organización, mientras que las oportunidades y amenazas son elementos externos a ésta. A continuación se presentan los resultados del análisis realizado.

Debilidades
<ul style="list-style-type: none">• Se podría perder cierto grado de confidencialidad en algunos datos que administra la organización.• Se permitiría el tránsito no controlado de datos propios de la organización hacia el exterior, por lo que sería complejo establecer responsabilidades para el control de los datos.• Se podría evidenciar la existencia de datos no convenientemente estructurados dentro de la organización, lo cual afectaría al uso adecuado de los mismos.
Amenazas
<ul style="list-style-type: none">• Se corre el riesgo de tener incursiones no deseadas en foros de discusión generados a partir de datos abiertos publicados.• Se corre el riesgo de brindar a la competencia datos que podrían ser considerados clave en el área de negocio en que se desenvuelve la organización.• Se podrían vislumbrar detalles que, si no existe un correcto análisis asociado, podrían verse como problemas existentes dentro de la organización.• Se podrían generar reticencias entre autoridades de la organización cuya mentalidad no va acorde con la apertura de datos de carácter organizacional.
Fortalezas
<ul style="list-style-type: none">• La organización se proyecta como un ente transparente ante la sociedad, lo cual le brinda mayor credibilidad ante sus clientes, personal u otras personas u organismos.• Se posiciona positivamente la reputación de la organización con respecto a sus competidores o fiscalizadores, lo cual le permite marcar diferencia en su segmento de mercado.

²⁷ http://www.agendadigital.gob.es/planes-actuaciones/Bibliotecaciudadesinteligentes/1.%20Plan/Plan_Nacional_de_Ciudades_Inteligentes.pdf

- Se proyecta la idea de que la organización cuenta con un **amplio respaldo documental** que sustenta todos los datos publicados.
- La publicación de datos abiertos supone un impacto económico para el tejido emprendedor que reutilice estos datos realizando productos o servicios de valor añadido.

Oportunidades

- Las personas vinculadas de alguna forma con la organización (clientes, colaboradores, aliados, etc.) pueden conocer con mayor detalle ciertos aspectos operativos o administrativos, lo cual podría generarle **retroalimentación objetiva** que es valiosa en la toma de decisiones.
- Se propicia una mejora sustancial en la **completitud** de los datos que emana de la organización, debido a los procesos de integración y depuración que deben llevarse a cabo para generar la apertura de datos.
- Se propicia una plataforma para mostrar que la organización cuenta con **altos estándares de calidad en sus datos**.

Tabla 1: Análisis DAFO del uso de datos abiertos en una organización

3.2 Trabajos relacionados

A continuación se presenta una recopilación de aportes relevantes relacionados con los conceptos que abarca este trabajo de tesis, con el objetivo de mostrar un panorama sobre los avances que se tienen en este campo de investigación.

3.2.1 Herramientas para publicación de datos abiertos

Como respuesta al aumento en la publicación de portales de datos abiertos, han venido aumentando también las alternativas tecnológicas que permiten a las distintas organizaciones publicar sus datos de manera ágil, simple y versátil. Existen herramientas diseñadas originalmente para la implementación y configuración de portales tradicionales de contenido que han sido adaptadas o al menos incluidas dentro de programas específicos relacionados con datos abiertos, y existen herramientas que han sido desarrolladas justamente para el diseño e implementación de portales de datos abiertos.

Un ejemplo de una herramienta que está siendo impulsada para estos fines y que tradicionalmente no estaba relacionada con datos abiertos es Microsoft SharePoint, herramienta altamente empleada a nivel mundial por todo tipo de organizaciones (se puede decir que actualmente es uno de los productos estrella de Microsoft Corporation). Reciente Microsoft lanzó el programa denominado CityNext²⁸, que busca impulsar la utilización de SharePoint y de otros productos complementarios (como SQLServer y Azure) para el desarrollo de portales de datos abiertos dentro de la filosofía del programa CityNext.

Como se mencionó, existen también herramientas precisamente diseñadas para implementar portales de datos abiertos, tales como Socrata²⁹ y Junar³⁰ (ampliamente utilizada en Costa Rica). Estas

²⁸ Microsoft CityNext: <http://www.microsoft.com/en-us/citynext/default.aspx>

²⁹ Socrata: <https://opendata.socrata.com>

³⁰ Junar: <http://www.junar.com>

herramientas proveen múltiples elementos y componentes nativos para la publicación de conjuntos de datos, para la integración y exportación de los mismos, y para muchas otras funcionalidades relacionadas con el manejo de los datos dentro del portal. Inclusive, algunas de estas herramientas facilitan el que los consumidores de datos abiertos puedan hacer uso de una API³¹ para reutilizarlos de manera automática, accediéndolos directamente desde otras aplicaciones que desarrollen.

Por su parte, algunos de los buscadores tradicionales de Internet han incorporado dentro de sus herramientas generadores de información proveniente de datos públicos, los cuales son recopilados de fuentes de datos también catalogados como públicas. Un ejemplo de ello es el reconocido buscador Google, el cual contiene una herramienta llamada Public Data³² que permite consultar, comparar e incluso diagramar información que en otros momentos pudo haber sido catalogada como privada. Esta herramienta incluye el lenguaje de Google llamado DSPL (*Developers Dataset Publishing Language*³³), el cual utiliza como parte de su plataforma para la carga, integración y publicación de datos públicos. A continuación se presenta un ejemplo que muestra un gráfico generado en Google – a partir de datos públicos – sobre la deuda de los gobiernos de países europeos, los cuales son datos provenientes de Eurostat, uno de los sitios de datos públicos más visitados en el mundo.

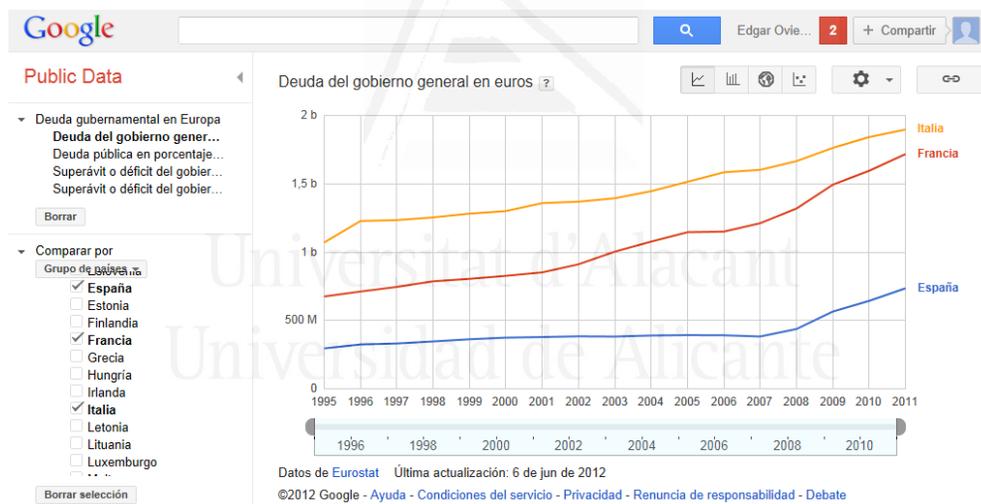


Figura 5: Gráfico generado en Google DSPL a partir de datos públicos

Las herramientas como el DSPL de Google no proveen mecanismos para la consideración de criterios de calidad más allá de aquellos relacionados con la integración de los mismos, es decir aquellos que están relacionados con que los datos sean exactos, completos o consistentes a través de diversas técnicas como por ejemplo, la detección y eliminación de duplicados (Rahm & Do 2000, Ananthakrishna et al 2002, Elmagarmid et al 2007).

³¹ API es la abreviatura de Interfaz de Programación de Aplicaciones. Es un conjunto de rutinas, protocolos y herramientas creadas para la creación de aplicaciones de software.

³² <http://www.google.com/publicdata>

³³ <https://developers.google.com/public-data/overview>



Por esto, a pesar de la existencia de herramientas cada vez más sofisticadas para la publicación y administración de datos abiertos, sigue siendo necesario el establecimiento de criterios de calidad evaluables para garantizar portales de datos abiertos de calidad.

3.2.2 Modelos de calidad para datos abiertos

Teniendo claro el concepto de un portal de datos abiertos, es evidente que la calidad de los datos recopilados determinará en un alto porcentaje la calidad de los resultados obtenidos al reutilizar dichos datos a través del desarrollo de servicios o productos de valor añadido. Por ello, resulta muy importante contar con criterios, simples pero efectivos, que permitan evaluar la calidad de los datos que se publican en un portal de datos abiertos, así como del portal en sí mismo. Por ejemplo, en (Heise & Naumann 2012), los autores destacan cómo la heterogeneidad de los datos que normalmente se incorporan al *open government* dificulta la posibilidad de hacer un análisis e integración de los datos acorde con las expectativas de los ciudadanos, lo cual limita en algún grado la proyección de transparencia que se busca. En dicho trabajo se presentan nuevas técnicas para realizar una mejor integración de datos públicos heterogéneos. Por lo tanto se tienen en cuenta nuevamente criterios de calidad relacionados con la integración de datos, pero se deja a un lado la consideración de otros criterios de calidad necesarios para satisfacer a la ciudadanía a la hora de analizar datos públicos en un entorno de gobierno abierto, como pudieran ser la reputación o la capacidad de visualización de los datos.

Por tanto, como parte de esta investigación se analizó el aporte realizado previamente por diversos autores sobre calidad de datos. En la actualidad existe no solamente una gran diversidad de técnicas para medir la calidad de los datos, sino también una serie de metodologías sugeridas para llevar a cabo dicha labor. Estas metodologías van mucho más allá que las simples técnicas de calidad de datos, pues consideran una serie de pasos previos y posteriores a la aplicación de las técnicas en sí, así como factores más allá de lo meramente técnico que permiten incluso poder compararlas entre ellas. Incluso, es posible combinar distintas técnicas en una misma metodología para garantizar procesos más completos de verificación de la calidad de datos.

En (Batini et al., 2009) se destaca el hecho que existen dos tipos de técnicas de calidad de datos:

a) Técnicas de calidad de datos basadas en los propios datos:

- Adquisición de nuevos datos que sustituyan aquellos valores inconsistentes.
- Estandarización o normalización, que consiste en remplazar o complementar aquellos valores que estén fuera de algún estándar, por ejemplo el remplazo de abreviaturas.
- Asociación de registros, que establece una relación entre datos de diferentes fuentes que se refieren a un mismo objeto o concepto.
- Integración de datos y esquemas, que consiste en definir una vista unificada de los datos proveídos por fuentes heterogéneas de datos.
- Confiabilidad de las fuentes, la cual consiste en seleccionar o discriminar las fuentes de datos con base en la calidad de sus datos.

- Localización y corrección de errores, que consiste en identificar y eliminar inconsistencias en los datos, detectando los registros que no cumplan con ciertas reglas de calidad preestablecidas.
 - Optimización de costo, que define acciones de mejora de calidad a través de un conjunto de dimensiones minimizando costos.
- b) Técnicas de calidad de datos basadas en los procesos que se aplican a los datos
- Control del proceso, que establece procedimientos de chequeo y control en el proceso de producción de datos, cuando: (1) se crean nuevos datos, (2) se actualizan conjuntos de datos, o (3) nuevos conjuntos de datos son accedidos por el proceso.
 - Rediseño del proceso, que se aplica para remover las causas de la baja calidad de los datos.

Partiendo de la clasificación anterior, los criterios o técnicas más empleados en modelos de calidad pertenecen a la categoría de “Técnicas de calidad de datos basadas en los propios datos”, pues coinciden en la relevancia de aspectos tales como confiabilidad de las fuentes, integración de datos y esquemas, y unificación de los registros.

Sin embargo, aunque muchos criterios de calidad de datos han sido ampliamente estudiados y pueden ser aplicados tanto a datos abiertos como a cualquier otro tipo de datos, existen criterios de calidad de datos específicos que deben considerarse para la evaluación de los datos abiertos, debido a que estos datos están disponibles para cualquiera, y por ende cualquier persona puede reutilizarlos.

El primer paso dado en este sentido es el modelo de calidad para datos abiertos propuesto por Tim Berners-Lee en mayo de 2010, durante el evento Gov 2.0 Expo 2010³⁴ y en un documento de la W3C³⁵, llamado Esquema (o modelo) 5 Estrellas³⁶. Este modelo establece cinco niveles de calidad etiquetados con estrellas según su la facilidad de uso por parte de los consumidores de datos:

- ★ Los datos están disponibles en la Web, independientemente del formato utilizado.
- ★★ Los datos se publican en la Web en un formato estructurado.
- ★★★ Los datos están publicados bajo un formato no propietario.
- ★★★★ Los datos se identifican mediante URLs de manera que sean fácilmente interpretables.
- ★★★★★ Los datos están vinculados con otros datos de manera que se encuentran contextualizados.

Tabla 2: Esquema del Modelo de 5 Estrellas

A continuación se presenta una representación gráfica que muestra de manera muy simple el sentido o la esencia de cada uno de los niveles de este Modelo 5 Estrellas.

³⁴ <http://www.youtube.com/watch?v=ga1aSJXCFe0>

³⁵ <http://www.w3.org/DesignIssues/LinkedData.html>

³⁶ <http://5stardata.info>

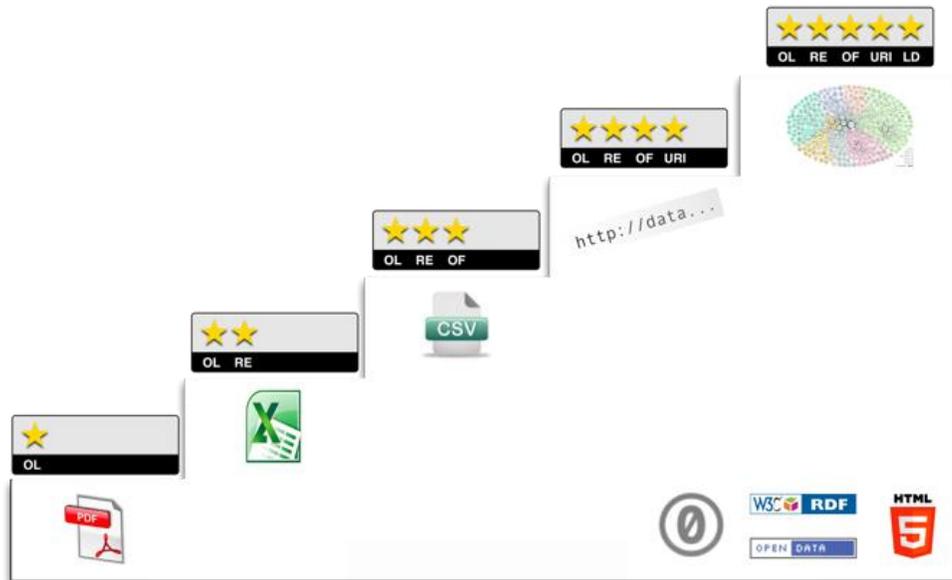


Figura 6: Representación gráfica del Modelo 5 Estrellas

Seguidamente se muestra un ejemplo del cumplimiento de cada uno de estos cinco niveles de calidad, tomando como base el portal de datos abiertos de la Municipalidad de Pérez Zeledón (explicado con mayor amplitud en el capítulo 5).

1ª Estrella: La información es publicada, independientemente del formato utilizado. En este caso se publica como una tabla dentro del propio portal.

Provincia, cantón y distrito	Total Rural y Urbano	Hombres	Mujeres
Pérez Zeledón	134534	65389	69145
San Isidro de El General	45327	21358	23969
El General	6373	3137	3236
Daniel Flores	33537	16148	17389
Rivas	6591	3306	3285
San Pedro	9102	4487	4615
Platanares	7203	3665	3538
Pejibaye	7995	4038	3957
Cajón	8542	4229	4313
Barú	2393	1212	1181
Río Nuevo	3061	1574	1487
Páramo	4410	2235	2175

Figura 7: Cumplimiento de la 1ª estrella

2ª Estrella: La información es accesible en un formato estructurado. En este caso, los datos son exportables en formato Excel.

Provincia, cantón y distrito	Total Rural	Hombres	Mujeres	Total Urbi	Hombres	Mujeres	Total Rural	Hombres	Mujeres
1 Provincia, cantón y distrito	134534	65389	69145	65255	30818	34437	69279	34571	34708
2 Pérez Zafedón	45327	21358	23969	33820	15747	18073	11507	5611	5896
3 San Isidro de El General	6373	3137	3236	-	-	-	6373	3137	3236
4 El General	33537	16148	17389	28693	13755	14938	4844	2393	2451
5 Daniel Flores	6591	3306	3285	-	-	-	6591	3306	3285
6 Rivas	9102	4487	4615	-	-	-	9102	4487	4615
7 San Pedro	7203	3665	3538	-	-	-	7203	3665	3538
8 Platanares	7995	4038	3957	1053	520	533	6942	3518	3424
9 Pejibayé	8542	4229	4313	1689	796	893	6853	3433	3420
10 Cajón	2393	1212	1181	-	-	-	2393	1212	1181
11 Barú	3061	1574	1487	-	-	-	3061	1574	1487
12 Río Nuevo	4410	2235	2175	-	-	-	4410	2235	2175
13 Páramo									

Figura 8: Cumplimiento de la 2ª estrella

3ª Estrella: La información es accesible en un formato estructurado no propietario. En este caso, los datos son exportables en formato CSV.

Provincia, cantón y distrito,"Total Rural y Urbano","Hombres","Mujeres","Total Urbano","Hombres","Mujeres","Total Rural","Hombres","Mujeres","Total"										
1 Provincia,cantón y distrito,"Total Rural y Urbano","Hombres","Mujeres","Total Urbano","Hombres","Mujeres","Total Rural","Hombres","Mujeres","Total"	"134534"	"65389"	"69145"	"65255"	"30818"	"34437"	"69279"	"34571"	"34708"	""
2 PÁrez Zafedón,"134534","65389","69145","65255","30818","34437","69279","34571","34708"	"45327"	"21358"	"23969"	"33820"	"15747"	"18073"	"11507"	"5611"	"5896"	""
3 San Isidro de El General,"45327","21358","23969","33820","15747","18073","11507","5611","5896"	"6373"	"3137"	"3236"	"-"	"-"	"-"	"6373"	"3137"	"3236"	""
4 El General,"6373","3137","3236","-","-","-","6373","3137","3236"	"33537"	"16148"	"17389"	"28693"	"13755"	"14938"	"4844"	"2393"	"2451"	""
5 Daniel Flores,"33537","16148","17389","28693","13755","14938","4844","2393","2451"	"6591"	"3306"	"3285"	"-"	"-"	"-"	"6591"	"3306"	"3285"	""
6 Rivas,"6591","3306","3285","-","-","-","6591","3306","3285"	"9102"	"4487"	"4615"	"-"	"-"	"-"	"9102"	"4487"	"4615"	""
7 San Pedro,"9102","4487","4615","-","-","-","9102","4487","4615"	"7203"	"3665"	"3538"	"-"	"-"	"-"	"7203"	"3665"	"3538"	""
8 Platanares,"7203","3665","3538","-","-","-","7203","3665","3538"	"7995"	"4038"	"3957"	"1053"	"520"	"533"	"6942"	"3518"	"3424"	""
9 Pejibayé,"7995","4038","3957","1053","520","533","6942","3518","3424"	"8542"	"4229"	"4313"	"1689"	"796"	"893"	"6853"	"3433"	"3420"	""
10 Cajón,"8542","4229","4313","1689","796","893","6853","3433","3420"	"2393"	"1212"	"1181"	"-"	"-"	"-"	"2393"	"1212"	"1181"	""
11 Barú,"2393","1212","1181","-","-","-","2393","1212","1181"	"3061"	"1574"	"1487"	"-"	"-"	"-"	"3061"	"1574"	"1487"	""
12 RÁ-o Nuevo,"3061","1574","1487","-","-","-","3061","1574","1487"	"4410"	"2235"	"2175"	"-"	"-"	"-"	"4410"	"2235"	"2175"	""
13 PÁiramo,"4410","2235","2175","-","-","-","4410","2235","2175"										

Figura 9: Cumplimiento de la 3ª estrella

4ª Estrella: La información es accesible directamente a través de un URL específico, el cual es generado a través de una de las opciones de la plataforma.



Figura 10: Cumplimiento de la 4ª estrella

5ª Estrella: El portal de datos abiertos tomado para el ejemplo no cumple con la quinta estrella; sin embargo, para ejemplificar cómo sí la cumpliría se podría suponer que existe un hipervínculo que redirige al sitio del Instituto Nacional de Estadísticas y Censos de Costa Rica (INEC³⁷) y otro al sitio de Wikipedia donde detalla información sobre Pérez Zeledón³⁸, los cuales se muestran a continuación:

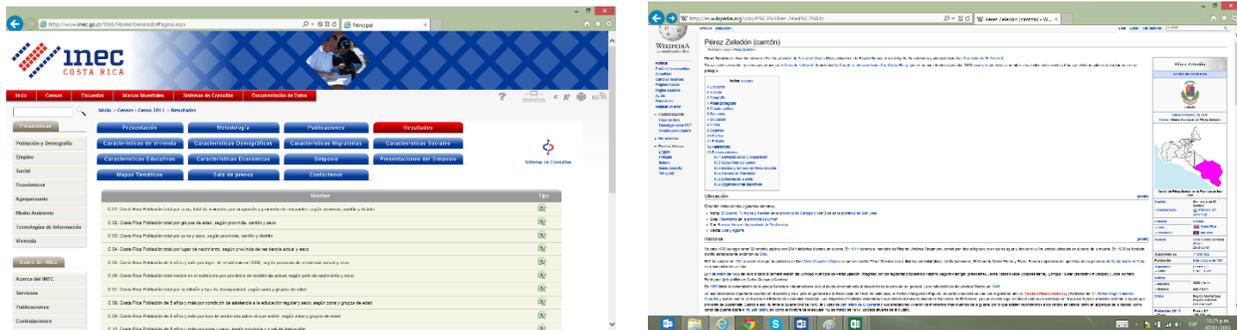


Figura 11: Cumplimiento de la 5ª estrella

Este simple pero difundido modelo de calidad conocido como Modelo 5 Estrellas presenta algunas carencias principalmente debido a que se centra únicamente en los dos aspectos de calidad de datos que se deben maximizar para que un dato sea abierto (tal y como se comentará en el siguiente capítulo).

Existen unos pocos trabajos de investigación (y, además, muy recientes) que intentan mejorar esta situación en relación a la falta de propuestas de calidad de datos abiertos, aunque la mayoría se centra en la evaluación de la calidad de los datos publicados en portales de datos abiertos, sin indagar en la carencia existente al no contar con modelos de calidad de datos propios para datos abiertos. Por ejemplo, en (Umbrich et al 2015a) y (Umbrich et al 2015b) se destaca que, a pesar de que las iniciativas de datos abiertos han gozado de gran popularidad, no ha habido trabajos de investigación que analizaran la calidad de los datos abiertos publicados. De hecho, en estos artículos se describe un procedimiento para medir la calidad de los datos abiertos existentes en determinados portales (en concreto los que usan la plataforma CKAN³⁹), sentando las bases de una monitorización de la calidad en el tiempo de 82 portales que contienen más de 160.000 conjuntos de datos. Por otra parte, en (Höchtl 2015) afirma que las propuestas de calidad de datos se deben centrar en mejorar la usabilidad de los portales de datos abiertos, centrándose en evitar enlaces "rotos", mejorar descripciones inexistentes o pobres, arreglar codificaciones erróneas o enmendar ficheros irregulares (por ejemplo, CSV mal formados).

3.2.3 Modelos de madurez para portales de datos abiertos

Algunos trabajos previos como el realizado en (Abella et al, 2014), hacen ver que “la disponibilidad de mecanismos de evaluación del uso y valor de la información publicada en portales de datos abiertos es limitada”, y ante ello proponen un modelo que se enfoca principalmente en la reutilización de los datos,

³⁷ <http://www.inec.go.cr>

³⁸ http://es.wikipedia.org/wiki/Perez_Zeledon

³⁹ <http://ckan.org/>



por considerar esto como el principal objetivo de la publicación de datos abiertos. El modelo, llamado Meloda, tiene su justificación en la falta de homogeneidad en los conjuntos de datos que se publican en este tipo de portales, y para determinar eso establece cuatro dimensiones que son:

1. Estándares o estructura técnica en que se ofrece el dato.
2. Acceso a la información o mecanismo por el cual se hace posible la descarga o conexión con la información.
3. Marco legal: licencia que se asigna al conjunto de datos.
4. Modelos de datos: modelo utilizado para publicar la información.

El modelo propone una forma cuantitativa para evaluar la reutilización de datos, dando un valor a cada dimensión y calculando la puntuación final como la raíz cuarta del producto de las puntuaciones obtenidas en cada dimensión. La razón por la que se considera insuficiente este modelo es porque, como antes se mencionó, se enfoca principalmente en la reutilización de los datos, dejando de lado otros aspectos previamente considerados como relevantes en la evaluación de la madurez de un portal de datos abiertos.

El segundo trabajo relacionado es el modelo de madurez propuesto en 2015 por el Open Data Institute del Reino Unido, llamado Open Data Maturity Model, el cual se basa en cinco temas y cinco niveles de progreso. Cada tema representa un área de operaciones de una organización y se divide en las áreas de actividad:

1. Procesos de administración de datos.
2. Conocimiento y habilidades.
3. Soporte y vinculación del cliente.
4. Inversión y desempeño financiero.
5. Supervisión estratégica.

Al ser analizado se determinó que la aplicación de este modelo resulta una labor altamente compleja, debido a las múltiples perspectivas (que combinan subdivisiones o subcategorías) con que pretende medir la madurez de un portal de datos abiertos.

El tercer trabajo tomado como referencia consiste en un modelo de madurez propuesto por la Organización Universitaria Interamericana de Chile, llamado OD-MM. Este modelo, de considerable complejidad de aplicación, tiene una estructura jerárquica de tres niveles, llamados dominios, subdominios y variables críticas. Se basa en las mejores prácticas internacionales de datos abiertos, que se recogen para ser incorporadas en 3 dominios del modelo:

- Perspectiva Institucional y Legal
- Perspectiva Tecnológica
- Perspectiva Ciudadana y Empresarial

Cada dominio a su vez se compone de 3 sub-dominios, siendo 9 subdominios en total:

1. Estrategias, Liderazgo e Institucionalidad
2. Leyes y Normas
3. Gestión
4. Seguridad y Disponibilidad
5. Acceso
6. Calidad de los Datos
7. Reutilización de los Datos
8. Desarrolladores
9. Participación y Colaboración

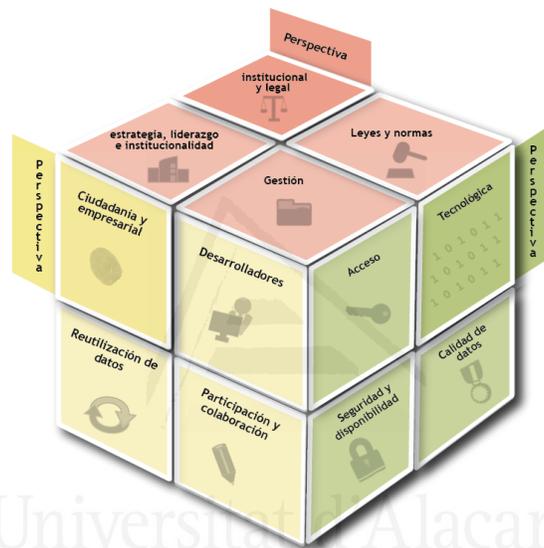


Figura 12: Representación del modelo de madurez de datos abiertos OD-MM



Universitat d'Alacant
Universidad de Alicante



Capítulo 4: Modelo de Madurez para Portales de Datos Abiertos

En este capítulo se presenta formalmente la propuesta de un Modelo de Madurez para Portales de Datos Abiertos, el cual se basa en unos criterios de calidad de datos abiertos definidos también en este trabajo de investigación.

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

Esta investigación fue desarrollada básicamente en dos etapas cuyos resultados son productos diferentes pero completamente entrelazados y con el mismo objetivo final: crear un modelo de madurez para portales de datos abiertos. En la primera etapa se planteó (y se obtuvo) como resultado el establecimiento de ciertas medidas o criterios de calidad que, de manera conjunta y con ciertas reglas establecidas para dicho fin, conformaron la base del modelo de calidad propuesto en (Oviedo et al 2013). Este modelo se explica detalladamente en la sección 4.1. En la segunda etapa de esta investigación se definió formalmente un modelo de madurez basado en los criterios del modelo de calidad (Oviedo et al 2015), el cual fue desarrollado utilizando el estándar genérico de ISO para modelos de madurez presentado en (Pérez-Mergarejo 2014). Este modelo de madurez para portales de datos abiertos se describe detalladamente en la sección 4.2.

4.1 Modelo de Calidad para Portales de Datos Abiertos

Como se ha mencionado con anterioridad, la calidad de los datos determina en un alto porcentaje la calidad de los resultados obtenidos al consumir dichos datos. A partir de esta premisa, es importante determinar cómo se puede evaluar la calidad de los datos publicados en un portal de datos abiertos, para garantizar que los resultados que surjan de su consumo sean de la calidad requerida. Además, es preciso desarrollar nuevos criterios de calidad de datos, ya que la calidad de datos abiertos difiere de la calidad de datos “tradicional” debido al origen de los datos: los datos abiertos no se encuentran en un entorno controlado, como sí se estaría en un escenario controlado si se consideran únicamente datos privados de una organización. Además, los datos abiertos están disponibles a todo el mundo y todo el mundo puede reutilizarlos (y, por ejemplo, generar nuevos datos abiertos), hecho este a tener en cuenta al definir criterios de calidad para datos abiertos. De esta manera, intuitivamente se puede pensar en la existencia de criterios de calidad preferentes para datos abiertos tales como la accesibilidad o la reputación de sus fuentes.

El punto de partida del modelo de calidad definido es el esquema 5 estrellas propuesto por Tim Berners-Lee (definido anteriormente en este documento). Al analizar el requerimiento que implica cada uno de los cinco niveles, se puede observar que todos los niveles están contenidos en dos criterios genéricos: disponibilidad de los datos y capacidad de reutilización de los datos, tal y como se muestra a continuación:

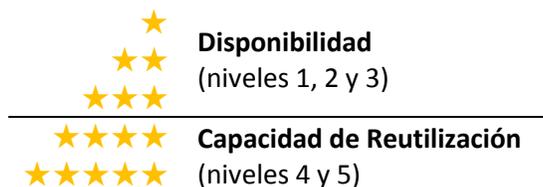


Figura 13: Equivalencia de los niveles del Esquema de 5 Estrellas

No obstante, existen otros criterios de calidad que se deben considerar partiendo de la base del concepto anteriormente explicado (en la sección 1.4) de adecuación al uso. En este esquema de 5 estrellas se deja de lado, por ejemplo, el concepto de integración de datos, el cual juega un papel

sumamente importante pues en la inmensa mayoría de los casos las fuentes de datos que alimentan un portal de datos abiertos son empleadas en origen para cumplir funciones propias de la operativa diaria de la organización, o sea, los datos no estaban preparados para ser consumidos fuera de su contexto original. Sobre la importancia de la integración existen varios trabajos que se centran en la definición del concepto de *dataspace* (Franklin et al 2005). Un *dataspace* ofrece una gestión de datos bajo demanda incluyendo integración de bases de datos heterogéneas, así como propuestas para analizar grandes colecciones de tablas de documentos HTML (Cafarella et al 2008). Resulta imprescindible, por tanto, que los procesos de integración estén bien definidos y que se apliquen periódicamente, de modo que se garantice que los datos publicados siempre estén integrados y representen fehacientemente la realidad.

Cabe destacar que muchos criterios de calidad de datos han sido ampliamente estudiados y pueden ser aplicados tanto a datos abiertos como a cualquier otro tipo de datos (Batini et al 2009). No obstante, el modelo de calidad propuesto en esta tesis plantea un modelo cualitativo que parte del esquema 5 estrellas y lo amplía teniendo en cuenta el propio concepto de “apertura” en los datos y considerando, no solamente la calidad de los conjuntos de datos abiertos, sino también la calidad de los portales donde se publican estos datos abiertos. La base del modelo de calidad de datos abiertos propuesto consiste en los siguientes criterios:

1. **Disponibilidad de los datos publicados**, es decir, el grado de dificultad requerido para poder acceder a los datos contenidos en el portal de datos abiertos. Por ejemplo, se considerará que cumple este criterio aquel portal de datos abiertos en el cual al ingresar el usuario pueda disponer de los datos que busca con suma facilidad, sin necesidad de navegar exhaustivamente.
2. **Capacidad de reutilización de los datos publicados**, es decir, el grado en que los datos contenidos en el portal de datos abiertos pueden reutilizarse para realizar productos y servicios de valor añadido, más allá de los que puede permitir el propio portal. Por ejemplo, si un portal de datos abiertos permite la exportación de conjuntos de datos se considerará que cumple con este criterio.
3. **Relevancia de los datos publicados**, entendida como el grado de frecuencia con que son consultados los datos publicados. Por ejemplo, un portal de datos abiertos cuyas estadísticas de acceso evidencien que es altamente consultado en su entorno, cumple este criterio.
4. **Reputación de los datos publicados**, es decir, el grado de credibilidad que tenga el portal de datos abiertos donde se encuentran los datos publicados. Por ejemplo, un portal de datos abiertos que se alimente de fuentes de datos de instituciones de gobierno se considerará de alta reputación.
5. **Granularidad de los datos publicados**, es decir, el nivel de detalle con que se publicaron los datos que se relaciona con el potencial de reutilización de dichos datos. Por ejemplo, en un portal de datos abiertos un municipio se espera que un conjunto de datos sobre algún tema demográfico pueda ser analizado a nivel de barrio o caserío para poder considerarlo de alta granularidad.
6. **Visualización de los datos publicados**, es decir, el grado de simplicidad y completitud con que se muestren gráficamente los datos a quien los accede. Por ejemplo, se considerará que un portal cumple con este criterio si muestra, para cada conjunto de datos, al menos un elemento gráfico que permita complementar su análisis.

Partiendo de estos criterios, se ha desarrollado un modelo para la evaluación de la calidad de los portales de datos abiertos, que incluye diferentes medidas de calidad que se consideran aplicables a datos abiertos. Estas nuevas medidas de calidad para datos abiertos fueron evaluadas y parametrizadas con el fin de poder ser incluidas en un modelo de madurez para la evaluación de un portal de datos abiertos. Concretamente, el modelo de calidad de datos abiertos desarrollado en (Oviedo et al 2013) consiste en un modelo cualitativo que comprende varias medidas de calidad aplicables a datos abiertos, considerando aspectos que no habían sido profundamente detallados en estudios o trabajos anteriores y que ameritaban ser incluidos dentro de un instrumento para la medición de la calidad de un portal de datos abiertos. A continuación se presenta el modelo de calidad para portales de datos abiertos.

4.1.1 *Medidas de calidad propuestas*

Como base del modelo de calidad propuesto, inicialmente se establecieron diferentes medidas de calidad que, de manera conjunta, permiten analizar la calidad tanto de los datos publicados como del portal empleado para su acceso. Dichas medidas se presentan a continuación, respaldadas por preguntas cuyas respuestas permiten definir el grado en que dicha medida o criterio se cumple en un portal de datos abiertos.

Disponibilidad de los datos publicados, es decir, el grado de dificultad para poder acceder a los datos.

- ✓ ¿Está la información publicada en formatos abiertos, de modo que para accederla no se requiere de herramientas de software específicas?
- ✓ ¿Está la información disponible para ser analizada al momento de observarla, sin necesidad de aplicarle algún proceso de transformación o conversión?
- ✓ ¿Está la información distribuida de manera que es comprensible y analizable sin necesidad de tener que acceder información adicional de otros sectores del portal o de otras fuentes?

Capacidad de reutilización de los datos publicados, es decir, el grado en que los datos pueden reutilizarse para realizar productos y servicios de valor añadido.

- ✓ ¿Es posible la exportación de toda (o parte de) la información publicada, para efectos de realizar análisis a dicha información de manera externa al portal?
- ✓ ¿Es acorde el nivel de detalle que permite el portal sobre cierto tema con respecto a la importancia o relevancia del mismo, o no se considera relevante este aspecto?
- ✓ ¿Incluye el portal herramientas para vincular entre sí temas que podrían resultar (para el consumidor de información) de alto valor el poder relacionarlos?

Relevancia de los datos publicados, es decir, la frecuencia con que se consultan los datos publicados.

- ✓ ¿Son consultados los datos con suficiente frecuencia como para ser considerados datos relevantes para los usuarios del portal dentro de su contexto?
- ✓ ¿Tienen todos los conjuntos de datos una frecuencia esperada o estimada de acceso que, de forma independiente, justifique su publicación?



- ✓ ¿Es referenciado el portal y sus conjuntos de datos notablemente por parte de otros sitios web?

Reputación de los datos publicados, es decir, el grado de credibilidad que tenga el portal donde se encuentran los datos publicados.

- ✓ ¿Se indica claramente cuáles son las fuentes de las cuales se obtuvo la información que está siendo publicada en el portal, de manera que estas fuentes puedan ser verificables?
- ✓ ¿Tienen prestigio o trayectoria dentro de su entorno las fuentes de las cuales proviene la información publicada en el portal?
- ✓ ¿Se respeta la fuente original o existe alguna manipulación que pueda poner en duda el nivel de credibilidad de los datos publicados?

Granularidad de los datos publicados, es decir, el nivel de detalle con que se publicaron los datos que se relaciona con el potencial de reutilización de dichos datos.

- ✓ ¿Se muestra la información con un nivel de detalle suficiente como para propiciar que pueda ser ampliamente reutilizada en multitud de escenarios?
- ✓ ¿Existen herramientas de ayuda dentro del portal que permitan ahondar en detalles más específicos de los publicados originalmente en el portal?
- ✓ ¿Es posible explorar a fondo en cualquier aspecto que se desee dentro de los temas que se abarcan en el portal (incluyendo toda la información referente a los metadatos), está limitado a algunos o no es posible del todo?

Visualización de los datos publicados, es decir, la capacidad de acceder a los datos gráficamente.

- ✓ ¿Utiliza el portal herramientas visuales acordes con el tipo de información que despliega, de modo que permita en el acto el análisis y evaluación de la misma?
- ✓ ¿Permite el portal utilizar distintas formas para visualizar la misma información o se limita a presentar cierta información de una única manera?
- ✓ ¿Aprovecha el portal al máximo todas las bondades que brinda una herramienta de este tipo de modo que se explota al máximo la información, o se cuenta con poco nivel de detalle?

A continuación se muestra un esquema con la definición de cada uno de los seis criterios originalmente planteados en el modelo de calidad:

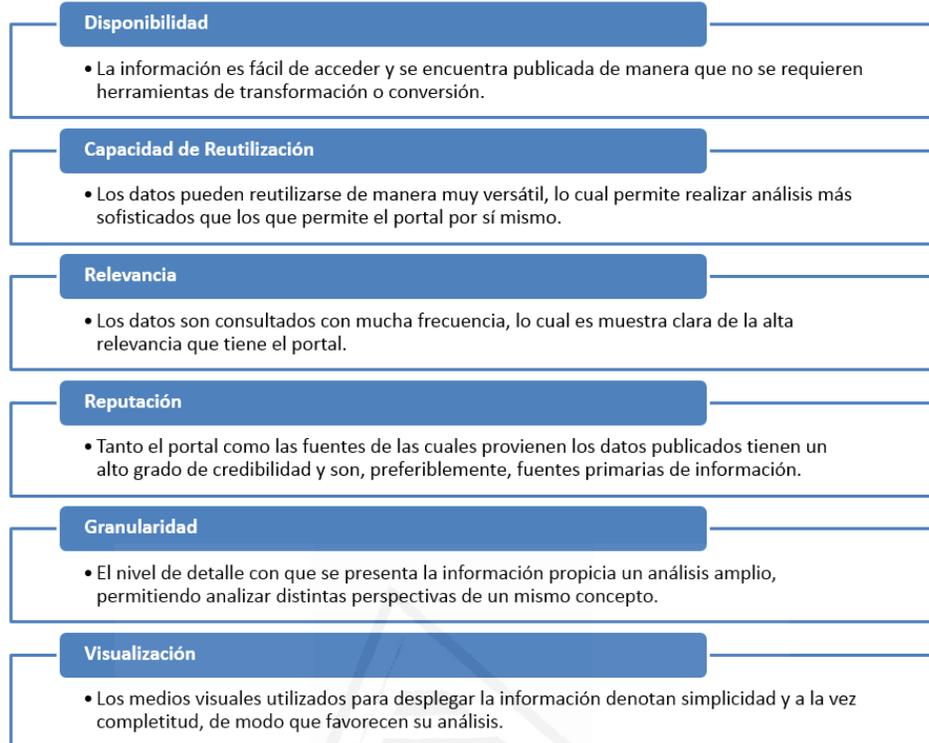


Figura 14: Criterios del modelo de calidad propuesto en (Oviedo et al 2013)

4.1.2 Factor de ajuste

Con el fin de ajustar adecuadamente la evaluación de la calidad de un portal de datos abiertos, tomando en consideración la criticidad del producto o servicio de valor añadido que consume la información que en éste se publiquen, los criterios que conforman el modelo de calidad propuesto se deben combinar con una variable adicional que tome en cuenta lo crítico o determinante que puede ser el uso que se prevé para los datos que fueron publicados, o bien, lo crítico o determinante del contexto dentro del cual se consumen los datos. Esto se fundamenta en que deben tener mayor peso ciertos criterios de calidad de datos si se trata de consumir, por ejemplo, los datos para realizar un análisis en un entorno financiero o bursátil, en comparación con un análisis llevado a cabo por parte de un ciudadano promedio, como podría ser analizar las tendencias de moda o los sitios de diversión más visitados actualmente. Es decir, no tiene la misma criticidad analizar datos abiertos relacionados con información accionaria de una empresa que analizar los resultados de encuestas abiertas realizadas a cierta población sobre un tema de relativamente poco impacto.

En Colombia fue desarrollada la aplicación móvil de datos abiertos llamada Verifíquese⁴⁰, la cual, a partir de la lectura del código de barras de la cédula de identidad de un ciudadano de ese país, permite observar de manera consolidada información de diversa índole incluyendo antecedentes legales,

⁴⁰ <http://verifique.se>

infracciones de tránsito, información financiera, afiliaciones y más. Sin duda, esta aplicación requiere que la reputación de sus fuentes de datos sea la máxima posible, pues de ello depende su credibilidad.

Por tanto, el peso o grado de inherencia que tendrá cada uno de los cinco criterios que conforman el modelo de calidad propuesto dependerá del nivel de criticidad de la información publicada o de su contexto. Se definió la siguiente matriz de asignación de pesos para los criterios que conforman el modelo de calidad propuesto, cuyos valores están asignados con base en la experiencia y no en cálculos o fundamentos matemáticos, a partir del conocimiento del autor en la aplicación de modelos de madurez tales como el CMM y dada su experiencia en el sector de la industria del software. Como puede notarse, cuanto mayor sea la criticidad de la reutilización mayor peso se otorga a los criterios de capacidad de reutilización, relevancia, reputación y granularidad, mientras que la disponibilidad y la visualización tienen mayor peso en ambientes de baja criticidad.

Criterio \ Criticidad de la reutilización	Baja	Media	Alta
Disponibilidad	20%	15%	10%
Capacidad de Reutilización	10%	15%	20%
Relevancia	10%	15%	20%
Reputación	15%	20%	20%
Granularidad	15%	20%	20%
Visualización	30%	15%	10%

Tabla 3: Factor de ajuste aplicable a los criterios en el modelo de calidad

4.1.3 Rangos para la evaluación final

Una vez considerados y aplicados los factores de ajuste (o pesos) a cada uno de los criterios y obtenido un valor final de la evaluación, el nivel de calidad del portal de datos abiertos se ubicará en una de las siguientes categorías, dependiendo del rango en que quede el valor:

Nivel de calidad	Rango
Destacado	$V > 0.80$
Notable	$0.65 < V \leq 0.80$
Moderado	$0.50 < V \leq 0.65$
Deficiente	$0.30 < V \leq 0.50$
Deplorable	$V \leq 0.30$

Tabla 4: Categorías para la valoración final del nivel de calidad

4.1.4 Aplicación del modelo

Para la aplicación de este modelo cualitativo de calidad propuesto, se debe asignar una calificación intuitiva (en un rango de 0 a 1, donde 1 es la calificación máxima) a cada uno de las medidas del modelo propuesto, respondiendo las interrogantes que se explicaron anteriormente para cada criterio. Con ello sería posible determinar, de una manera cualitativa, el grado de calidad de los datos publicados en un

portal de datos abiertos mediante una lista de comprobación que pudiera formar parte de una encuesta a los propios usuarios.

A continuación se presenta un ejemplo de la aplicación del modelo de calidad propuesto a un portal de datos abiertos ficticio, cuyo contexto se considera en primera instancia de alta criticidad y en segunda instancia de baja criticidad, lo cual repercute tanto en los pesos asignados como en el resultado final.

Ejemplo #1: En un contexto de criticidad ALTA

En este ejemplo los valores asignados a cada criterio son los siguientes:

Criterio	Peso	Calificación
Disponibilidad	10%	0.8
Capacidad de Reutilización	20%	0.8
Relevancia	20%	0.8
Reputación	20%	0.8
Granularidad	20%	1
Visualización	10%	0.6
Promedio ponderado:		0.82 DESTACADO

Tabla 5: Ejemplo de aplicación del modelo en un entorno de criticidad alta

Ejemplo #2: En un contexto de criticidad BAJA

En este ejemplo los valores asignados a cada criterio son los siguientes:

Criterio	Peso	Calificación
Disponibilidad	20%	0.8
Capacidad de Reutilización	10%	0.8
Relevancia	10%	0.8
Reputación	15%	0.8
Granularidad	15%	1
Visualización	30%	0.6
Promedio ponderado:		0.77 NOTABLE

Tabla 6: Ejemplo de aplicación del modelo en un entorno de criticidad baja

Como puede observarse en el ejemplo anterior, el peso que se asigna a cada criterio a partir de la criticidad del entorno del portal que se evalúa, tiene influencia directa en el cálculo de la calificación final, lo cual incluso implica que un portal con similares características pero en entornos con distintos niveles de criticidad, pueden ser considerados en distintos niveles de calidad según este modelo.

En el ejemplo, el portal sería evaluado como “destacado” si fuese dentro de un entorno de criticidad alta y como “notable” si fuese dentro de un entorno de criticidad baja. La principal razón de lo anterior



radica en el peso asignado al criterio de Visualización, partiendo de que alguien con mucha experiencia (entorno de alta criticidad) no precisa de mayor apoyo en cuanto a visualización mientras que alguien de poca experiencia (entorno de baja criticidad) sí requiere de mecanismos que le simplifiquen la visualización. Por ello, en un entorno de criticidad alta el peso de la Visualización es de tan sólo un 10% pero en un entorno de criticidad baja este criterio es trascendental por lo que tiene un peso de un 30% (siendo de hecho el de mayor peso). Consecuentemente, en un entorno de alta criticidad existen otros criterios mucho más relevantes, por ejemplo la Capacidad de Reutilización cuyo peso (20%) es el doble con respecto al que tiene en un entorno de baja criticidad (10%).

4.2 Modelos de madurez

Además de la necesidad cada vez más frecuente y justificada de evaluar las características de un proceso, de un producto o incluso de una política organizacional para determinar si cuenta con un grado satisfactorio de calidad, surge también el deseo de contar con modelos que permitan, a partir de la medición realizada, ubicar al ente evaluado en uno de varios niveles de manera que esté justificado decir que cuenta con ese nivel de madurez y a la vez sea posible compararlo con otros entes también valorados bajo ese mismo esquema.

En (Paulk 2009) se define modelo de madurez como un conjunto estructurado de elementos y requisitos que condicionan el nivel de madurez de un ente en un aspecto determinado, estableciendo para ello etapas o niveles de madurez en un orden claro, discreto y absoluto. Una de las principales características de un buen modelo de madurez es que debe sugerir de manera explícita la posible evolución del ente evaluado en dicho aspecto, marcando una especie de hoja de ruta para la mejora continua. De hecho, un modelo de madurez puede verse como un mapa que guía a una organización en la implementación de buenas prácticas, ofreciendo un punto de partida y describiendo un camino de mejoramiento evolutivo, desde los procesos inconsistentes hasta los más maduros de la organización. Permite evaluar el estado de desarrollo de una organización o proceso de negocio, trazar claramente estrategias de mejoras para alcanzar los objetivos previstos e identificar las áreas donde la organización debe enfocarse para mejorar.

Los modelos de madurez constituyen una evolución de las prácticas para gestionar la calidad. Fueron concebidos inicialmente para la industria del software y en la actualidad el área de aplicación es muy diversa. Se pueden encontrar aplicaciones en el desarrollo de software, en la gestión de proyectos, en la gestión del conocimiento, en el desarrollo de los procesos, en la gestión de procesos de negocio, en la administración de cadenas de suministros y otros.

En los últimos años se han venido empleando cada vez con más frecuencia los modelos de madurez, siendo quizá los más reconocidos el CMM (posteriormente llamado CMMi⁴¹) desarrollado por el SEI

⁴¹ www.sei.cmu.edu/cmmi

(Software Engineering Institute) y el BPMM⁴² de la OMG (Grupo de Administración de Objetos), publicado en 2008.

Como se muestra en (Pérez-Mergarejo 2014), en el año 2009 ISO estableció un esquema genérico para modelos de madurez, el cual se muestra a continuación:

Elemento clave	Nivel de madurez hacia el éxito sostenido			
	Nivel 1	Nivel 2	<...>	Nivel máximo
Elemento 1	Criterio 1 Nivel de base			Criterio 1 Mejor práctica
<...>				
Elemento N	Criterio N Nivel de base			Criterio N Mejor práctica

Tabla 7: Esquema genérico para modelos de madurez establecido por ISO

4.2.1 Historia del Modelo de Capacidad de Madurez (CMM)

En agosto de 1986 el Instituto de Ingeniería de Software (SEI) de la Universidad Carnegie Mellon con el apoyo del MITRE Corporation desarrolló un *framework* de madurez de procesos que ayudaría a las organizaciones a mejorar sus procesos de desarrollo de software. Este esfuerzo inició en respuesta a una solicitud del Gobierno Federal de los Estados Unidos por contar con un método para evaluar la capacidad de las empresas de software que le proveían servicios. En junio de 1987 el SEI lanzó una breve descripción del marco madurez de los procesos de software y, en septiembre de 1987, un cuestionario preliminar de madurez. Con base en la experiencia del uso del marco de madurez de los procesos de software y el cuestionario de madurez para el diagnóstico de problemas y la mejora de los procesos, el SEI formalizó los conceptos como el *Capability Maturity Model for Software* (Software CMM).

La versión 1.0 del modelo fue publicado en 1991 por Mark Paulk y la versión 1.1 fue lanzada en 1993 y publicada como libro en 1995, también por Paulk. Posteriormente, la liberación de una versión 2 del CMM Software fue suspendida para dar paso al modelo de Integración CMM (conocido como CMMI), pero inspiró a muchas otras normas y *frameworks*, incluyendo el People CMM, el CMM de Ingeniería de Sistemas y el CMM de Ingeniería de Seguridad de Sistemas. También, CMM fue uno de los pilares en el desarrollo de la norma ISO / IEC 15504 para la evaluación de procesos, sobre todo la parte 7 de evaluación de la madurez organizacional (ISO 2008).

4.2.2 ¿Cómo funciona el CMM?

En esencia, el CMM cuenta con 5 niveles, cada uno de los cuales representa un posible grado de madurez con que cuenta la organización en la ejecución de sus procesos de desarrollo de software. El nivel 1 representa el nivel Inicial y no contempla el cumplimiento de ningún requisito en particular. Del nivel 2 al nivel 5 se consideran diferentes áreas claves de proceso (o KPAs, luego llamadas sólo PAs), de

⁴² www.omg.org/spec/BPM

modo que al cumplir con todas las áreas de cierto nivel se dice que la organización cuenta con ese nivel de madurez. A continuación se presenta una imagen que sintetiza la idea general del CMM.

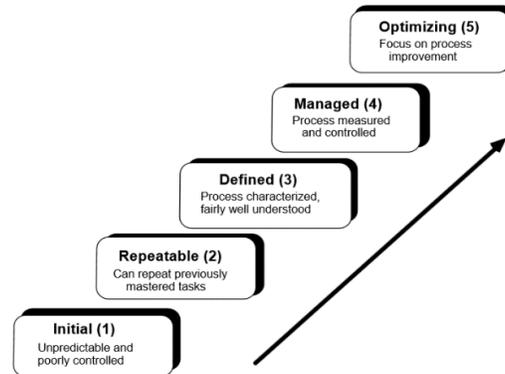


Figura 15: Niveles del Modelo de Capacidad de Madurez (CMM)

Los modelos de madurez de procesos categorizan o establecen niveles con base en prácticas ya implantadas (es decir, funcionando) y permiten a una organización saber dónde está para plantearse metas, con el fin de alcanzar niveles superiores en algunas áreas clave, lo cual logran mediante el uso de planes de mejora y mediante la ejecución y el seguimiento de los proyectos. No obstante, los modelos de madurez no solamente pueden ser empleados para medir la madurez de procesos, sino que se pueden adaptar para medir la madurez que refleja un producto.

En esta investigación se optó por utilizar el formato de un modelo de madurez al considerar que permite, de una manera simple e intuitiva, medir la madurez de un portal de datos abiertos a partir del cumplimiento de diversos criterios de calidad, cada uno representado a través de un requisito específico en cada uno de los niveles del modelo.

4.3 Base del modelo propuesto

El modelo de calidad de datos abiertos definido en (Oviedo et al 2013) provee los criterios que deben ser considerados para obtener una evaluación o grado de calidad para un portal de datos abiertos. Al realizar el modelo de madurez se evolucionó el modelo de calidad anteriormente definido con el fin de considerar conceptos importantes que merecían mayor atención dentro del modelo de madurez que se pretendía desarrollar. Además, se consideró que había criterios que podían combinarse en uno solo por las similitudes o afinidades que tenían. Producto de ese análisis se dio la evolución definida en la tabla 23, en cuanto a los criterios de calidad propuestos.

Criterios del Modelo de Calidad del 2013	Transformación	Criterios del Modelo de Madurez del 2015
Disponibilidad	se integran en	Disponibilidad y Reutilización
Capacidad de Reutilización		
Relevancia	se mantiene como	Relevancia
Reputación	se mantiene como	Reputación



Granularidad	se agrega el concepto de Integración	Granularidad e Integración
Visualización	se agrega el concepto de Usabilidad	Visualización y Usabilidad

Tabla 8: Evolución de los criterios del modelo de calidad al modelo de madurez

4.4 Niveles del Modelo de Madurez para Portales de Datos Abiertos

Atendiendo el esquema propuesto por ISO, el modelo de madurez para portales de datos abiertos que se propone como resultado de este trabajo de investigación está conformado por 4 niveles, donde el nivel 0 es el más bajo posible y el nivel 3 es el nivel más alto posible. A continuación se explica cada uno de los niveles del modelo.

4.4.1 Nivel 0: Incompleto

Se considerará que un portal de datos abiertos está en nivel 0 o Incompleto cuando no cumpla con todos los requisitos establecidos para el nivel 1, aun cuando esté cumpliendo algunos de dichos requisitos. Este nivel denota que el portal de datos abiertos evaluado no cumple ni siquiera con las características mínimas que se esperan en un recurso de este tipo.

4.4.2 Nivel 1: Básico

Un portal de datos abiertos podrá ubicarse en nivel 1 o Básico cuando cumpla con todos los requisitos establecidos para los diferentes criterios en el nivel 1, pero no cumpla plenamente con los requisitos establecidos para el nivel 2. Este nivel denota que el portal de datos abiertos evaluado satisface las características mínimas que se esperan en cada uno de los criterios de calidad que posee el modelo.

4.4.3 Nivel 2: Notable

El nivel 2 o Notable se asignará a aquellos portales de datos abiertos que cumplan con todos los requisitos establecidos para cada criterio en dicho nivel, lo cual denota que es un portal que posee características por encima del promedio esperado (que es el nivel 1 o Básico), pero no cumple con todos los requisitos establecidos para el nivel 3.

4.4.4 Nivel 3: Destacable

El nivel 3 o Destacable es el máximo nivel que puede recibir un portal de datos abiertos mediante la aplicación de este modelo de madurez. Lo obtendrán únicamente aquellos portales que cumplan con todos los requisitos del modelo, en todas las categorías de todos los niveles, lo cual evidentemente denota que se trata de un portal sumamente robusto que satisface las más altas expectativas de sus usuarios.

4.5 Criterios del Modelo de Madurez para Portales de Datos Abiertos

Como se anticipó en la sección 4.3, para este modelo se han establecido 5 criterios de calidad con base en los cuales será posible determinar el nivel de madurez de un portal de datos abiertos. Para cada criterio de calidad se ha establecido un requisito en cada uno de los niveles del modelo, tal y como se detalla a continuación:

4.5.1 Requisitos del criterio D/R – Disponibilidad y Reutilización

Los conceptos de disponibilidad y reutilización, que en el modelo de calidad definido en (Oviedo et al, 2013) aparecían como criterios independientes se presentan en este modelo de forma conjunta y pretenden evaluar en qué grado evoluciona el portal de datos abiertos en relación a su capacidad para hacer disponibles los datos y que estos sean reutilizados.

Se cumple el nivel 1 o Básico si solamente se muestra la información en un formato abierto, permitiendo su análisis y exportación de manera individual. El siguiente nivel se alcanza si permite integrar distintos conjuntos de datos y exportarlos de forma combinada. Finalmente, se dice que alcanza su máxima madurez en este criterio cuando el portal demuestre que su grado de reutilización es tal que existen comunidades que hacen uso de los datos abiertos publicados en él.

A continuación se presentan los requisitos de disponibilidad y reutilización que han sido establecidos para cada uno de los niveles de madurez del modelo.

Nivel	Resumen	Requisito
Nivel 0: Incompleto	N/A	N/A
Nivel 1: Básico	Se muestran conjuntos de datos disponibles y exportables independientemente.	- La información está publicada en un formato abierto, de modo que cada conjunto de datos puede ser analizado desde el propio portal y exportable de manera individual.
Nivel 2: Notable	Información integrable y exportable de forma combinada.	- El portal permite integrar distintos conjuntos de datos y otros elementos (como gráficos) para exportarlos de manera combinada con el fin de ser analizados desde otras herramientas.
Nivel 3: Destacable	Los datos del portal son empleados por comunidades dedicadas a la reutilización de datos.	- Existen comunidades dedicadas a la reutilización de datos que desarrollan o emplean aplicaciones informáticas que hacen uso de los datos abiertos publicados en el portal.

Tabla 9: Requisitos de Disponibilidad y Reutilización por cada nivel de madurez

4.5.2 Requisitos del criterio REL – Relevancia

La relevancia de un portal de datos abiertos se determinará en función de la frecuencia de accesos y consultas por parte de quienes requieran consumir los datos disponibles en dicho portal.

Se cumple el nivel 1 o Básico si se considera que el conjunto de datos que están publicados en el portal de forma conjunta abarcan un amplio segmento, suficientemente representativo, del entorno o ámbito al que se refiere el portal. Por ejemplo, si se tratara de un portal de datos abiertos de una organización encargada de la administración de las reservas forestales de un país, se cumpliría con el requisito del nivel básico si la información que está accesible en el portal incluye datos sobre la totalidad o al menos una mayoría significativa de las reservas forestales de ese país. El siguiente nivel se alcanza si el portal de datos abiertos es referenciado por otros sitios de consulta sobre el mismo tema al que se refiere dicho portal. Por último, se dice que un portal de datos abiertos alcanza su máximo nivel de madurez en este criterio cuando se considere a dicho portal como la fuente oficial de consulta de datos abiertos sobre el tema al que se refiere dicho portal.

A continuación se presentan los requisitos de relevancia que han sido establecidos para cada uno de los niveles de madurez del modelo.

Nivel	Resumen	Requisito
Nivel 0: Incompleto	N/A	N/A
Nivel 1: Básico	<i>Se muestra información suficientemente representativa de su entorno.</i>	- El portal contiene información que, considerando todos los conjuntos de datos publicados, representa una amplia porción de la realidad del entorno al que pertenece.
Nivel 2: Notable	<i>El portal puede considerarse altamente referenciado dentro de su entorno.</i>	- El portal es referenciado con frecuencia por parte de otros sitios de consulta, como para ser considerado una fuente relevante de consulta sobre el tema.
Nivel 3: Destacable	<i>El portal es el sitio oficial de datos abiertos en su ámbito local.</i>	- El portal es considerado como la fuente oficial de consulta de datos abiertos dentro de su entorno temático y geográfico.

Tabla 10: Requisitos de Relevancia por cada nivel de madurez

4.5.3 Requisitos del criterio REP – Reputación

La reputación de un portal de datos abiertos está en función de que se respeten las fuentes originales de los datos publicados, de lo fidedignos que estos sean y de que se actualicen con la frecuencia que se amerita dado su contexto.

El nivel 1 o Básico se obtiene si se considera que las fuentes de datos son indicadas debidamente en el portal y se respeta el contenido original, es decir, se garantiza que no existe manipulación a los datos originales. Para obtener el nivel 2 o Notable se evalúa el prestigio o la oficialidad con que cuenten las fuentes de datos que el portal emplea, es decir, se verifica que dichas fuentes de datos sean de carácter oficial, idealmente publicados o puestos a disposición por parte de organizaciones que sean referentes reconocidos en el mismo ámbito. El último nivel se alcanza en aquellos casos en los que se garantice una frecuencia de actualización tal que, considerando la criticidad y otras características del entorno, asegure que quien acceda y analice los datos siempre contará con información vigente al momento.

A continuación se presentan los requisitos de reputación que han sido establecidos para cada uno de los niveles de madurez del modelo.

Nivel	Resumen	Requisito
Nivel 0: Incompleto	N/A	N/A
Nivel 1: Básico	<i>Se indican las fuentes de datos y se respeta su contenido original</i>	- Las fuentes de datos que alimentan el portal son claramente indicadas y además se respetan los datos originales.
Nivel 2: Notable	<i>Los datos provienen de fuentes oficiales o reconocidas</i>	- Las fuentes de datos que alimentan el portal contienen información proveniente de fuentes oficiales o reconocidas en su ámbito, las cuales a su vez cuentan con proveedores respetables.
Nivel 3: Destacable	<i>Las fuentes de datos se actualizan de manera frecuente y conveniente</i>	- Las fuentes de datos que alimentan el portal se actualizan con una frecuencia adecuada según la temática y el entorno.

Tabla 11: Requisitos de Reputación por cada nivel de madurez

4.5.4 Requisitos del criterio G/I – Granularidad e Integración

El criterio de granularidad se refiere al nivel de detalle con que se muestran los datos en un portal de datos abiertos, es decir, en qué grado están siendo consideradas todas las distintas categorías que existen de cierto concepto dentro del entorno en que está siendo analizado. Por su parte, integración se refiere al grado en que el portal, como herramienta informática para el consumo de información, permite que mientras se esté analizando un conjunto de dato sea posible referenciar otros datos que existen dentro del portal o inclusive fuera de éste, con el fin de mejorar el análisis que se hace de los datos consultados.

El nivel 1 o Básico lo obtendrán aquellos portales cuyos conjuntos de datos muestren los datos en el mayor nivel de detalle que es posible para su entorno y su alcance. El nivel de madurez será de nivel 2 o Notable cuando, ante una consulta específica, el portal refiera a otros conjuntos de datos que tengan relación con el tema analizado, y que puedan representar información adicional de alto valor para quien está accediendo el portal. El máximo nivel de madurez en este criterio se obtiene cuando el portal cuente con la funcionalidad de permitir al usuario incorporar datos externos al portal, es decir, datos abiertos que residen en otras fuentes pero que podrían complementar el análisis que se está haciendo del conjunto de datos consultado.

A continuación se presentan los requisitos de granularidad e integración que han sido establecidos para cada uno de los niveles de madurez del modelo.

Nivel	Resumen	Requisito
Nivel 0: Incompleto	N/A	N/A
Nivel 1: Básico	Los datos se muestran con el máximo nivel de detalle posible	- La información se muestra con el máximo nivel de granularidad posible según el tema y el contexto de los datos publicados, considerando aspectos geográficos, demográficos y otros.
Nivel 2: Notable	El portal sugiere referencias internas sobre información complementaria	- Cuando se hace un análisis de un conjunto de datos, el portal sugiere algunas referencias a datos del mismo portal que tienen relación y que podrían complementar el análisis.
Nivel 3: Destacable	Es posible incluir información externa para complementar el análisis de un conjunto de datos	- El portal permite incorporar datos externos (publicados en otros sitios o portales) como parte del análisis que permite realizar de un conjunto de datos específico.

Tabla 12: Requisitos de Granularidad e Integración por cada nivel de madurez

4.5.5 Requisitos del criterio V/U – Visualización y Usabilidad

El quinto criterio que incluye este modelo de madurez es el de visualización y usabilidad, el cual pretende determinar las facilidades de interfaz que brinda el portal para consumir, entender y analizar los datos abiertos que están publicados en éste.

Se obtiene el nivel 1 o Básico cuando el portal de datos abiertos, como mínimo, permite analizar la información a través de elementos visuales como gráficos, histogramas o elementos similares. Cuando un portal de datos abiertos permita que la misma información sea graficada dinámicamente en distintos formatos, se considerará que éste cuenta con un nivel de madurez 2. Finalmente, el mayor nivel de

madurez se obtendrá cuando el portal, habiendo cumplido previamente con el nivel 2 o Notable, permita al usuario emplear aplicaciones o herramientas de análisis que hacen uso de las últimas tecnologías disponibles, como actualmente serían funcionalidades avanzadas de inteligencia de negocios o aplicaciones móviles.

A continuación se presentan los requisitos de visualización y usabilidad que han sido establecidos para cada uno de los niveles de madurez del modelo.

Nivel	Resumen	Requisito
Nivel 0: Incompleto	N/A	N/A
Nivel 1: Básico	<i>Se cuenta con herramientas para analizar la información usando gráficas</i>	- El portal facilita herramientas para analizar la información contenida en cada conjunto de datos, empleando para ello gráficas.
Nivel 2: Notable	<i>Se puede visualizar y analizar la información en diferentes formatos</i>	- El portal permite la visualización de la misma información en diferentes formatos.
Nivel 3: Destacable	<i>Se cuenta con aplicaciones de última generación para el análisis de datos</i>	- El portal posee aplicaciones o herramientas de consumo de datos que aprovechan las más modernas tecnologías disponibles en el mercado.

Tabla 13: Requisitos de Visualización y Usabilidad por cada nivel de madurez

4.6 Aplicación del Modelo de Madurez para Portales de Datos Abiertos

La mecánica de aplicación del modelo es sumamente sencilla y consiste en verificar – basado en la medición de los criterios de calidad a partir del juicio experto – el cumplimiento de cada uno de los requisitos de calidad que fueron establecidos para cada criterio en cada uno de los niveles, en cuyo caso el portal habrá alcanzado dicho nivel de madurez.

Dentro de un mismo criterio, al evaluar un portal de datos abiertos sólo se podrá cumplir con el requisito de un nivel específico si previamente cumplió con los requisitos de los niveles inferiores a dicho nivel.

En caso que al aplicar el modelo a un portal de datos abiertos se determine que éste cumple a cabalidad con cuatro de los cinco requisitos del nivel N y cumple parcialmente con el otro requisito de ese nivel, se podrá decir que el portal se encuentra en el nivel N –<criterio>, indicando luego del símbolo “–” el acrónimo de tres letras del criterio que sólo cumple parcialmente. Esto se hace con el objetivo que persigue justamente un modelo de madurez, que es plantear el reto de alcanzar el siguiente nivel.

Finalmente, en caso que un portal no alcance siquiera el nivel 1 –XXX, se considerará que el nivel de calidad y madurez de dicho portal es nivel 0 o Incompleto.

4.7 Modelo de Madurez para Portales de Datos Abiertos

A continuación se muestra el Modelo de Madurez para Portales de Datos Abiertos de forma completa.

Nivel	Criterio de Calidad	Requisito de Calidad
3 Destacable	D/R <i>Disponibilidad y Reutilización</i>	- Existen comunidades dedicadas a la reutilización de datos que desarrollan o emplean aplicaciones informáticas que hacen uso de los datos abiertos publicados en el portal.
	REL <i>Relevancia</i>	- El portal es considerado como la fuente oficial de consulta de datos abiertos dentro de su entorno temático y geográfico.
	REP <i>Reputación</i>	- Las fuentes de datos que alimentan el portal se actualizan con una frecuencia adecuada según la temática y el entorno.
	G/I <i>Granularidad e Integración</i>	- El portal permite incorporar datos externos (publicados en otros sitios o portales) como parte del análisis que permite realizar de un conjunto de datos específico.
	V/U <i>Visualización y Usabilidad</i>	- El portal posee aplicaciones o herramientas de análisis de datos que aprovechan las más modernas tecnologías disponibles en el mercado.
2 Notable	D/R <i>Disponibilidad y Reutilización</i>	- El portal permite integrar distintos conjuntos de datos y otros elementos (como gráficos) para exportarlos de manera combinada con el fin de ser analizados desde otras herramientas.
	REL <i>Relevancia</i>	- El portal es referenciado con suficiente frecuencia por parte de otros sitios de consulta, como para ser considerado una fuente relevante de consulta sobre el tema.
	REP <i>Reputación</i>	- Las fuentes de datos que alimentan el portal contienen información proveniente de fuentes oficiales o reconocidas en su ámbito, las cuales a su vez cuentan con proveedores respetables.
	G/I <i>Granularidad e Integración</i>	- Cuando se hace un análisis de un conjunto de datos, el portal muestra algunas referencias a datos del mismo portal que tienen relación y que podrían complementar el análisis.
	V/U <i>Visualización y Usabilidad</i>	- El portal permite la visualización de la misma información en diferentes formatos.
1 Básico	D/R <i>Disponibilidad y Reutilización</i>	- La información está publicada en un formato abierto, de modo que cada conjunto de datos puede ser analizado desde el propio portal y exportable de manera individual.
	REL <i>Relevancia</i>	- El portal contiene información que, considerando todos los conjuntos de datos publicados, representa una amplia porción de la realidad del entorno al que pertenece.
	REP <i>Reputación</i>	- Las fuentes de datos que alimentan el portal son claramente indicadas y además se respetan los datos originales.
	G/I <i>Granularidad e Integración</i>	- La información se muestra con el máximo nivel de granularidad posible según el tema y el contexto de los datos publicados, considerando aspectos geográficos, demográficos y otros.
	V/U <i>Visualización y Usabilidad</i>	- El portal facilita herramientas para consumir la información contenida en cada conjunto de datos, empleando para ello gráficas.
0 Incompleto		

Tabla 14: Modelo de Madurez para Portales de Datos Abiertos



Capítulo 5: Casos de Aplicación del Modelo de Madurez para Portales de Datos Abiertos

En este capítulo se muestran los resultados obtenidos de la aplicación del Modelo de Madurez para Portales de Datos Abiertos a tres portales que están actualmente en uso en Costa Rica:

- Portal de Datos Abiertos de la Municipalidad de Pérez Zeledón
- Portal de Datos Abiertos de la Municipalidad de Palmares
- Portal de Datos Abiertos del Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT)



Universitat d'Alacant
Universidad de Alicante

A continuación se presenta una descripción de los tres casos de estudio donde fue aplicado el modelo de madurez de portales de datos abiertos generado a partir de esta investigación, así como los resultados obtenidos de dicho proceso de evaluación.

5.1 Portal de Datos Abiertos de la Municipalidad de Pérez Zeledón

El primer caso de estudio incluido en este trabajo de investigación es el portal de datos abiertos de la Municipalidad de Pérez Zeledón⁴³, un cantón de la provincia de San José ubicado en la zona sur de Costa Rica, con una población cercana a los 140.000 habitantes.

Este portal de datos abiertos fue implementado en la plataforma JUNAR, la cual brinda una serie de facilidades orientadas justamente al desarrollo y publicación de portales de este tipo. En éste se publicó información de diversa índole, toda ella de carácter público, pero que tradicionalmente no había estado al alcance de la población (al menos no a través de la Web). Entre la información más relevante está:

- Censo de población.
- Distribución de ingresos y egresos.
- Salarios y otros beneficios de los funcionarios municipales.
- Endeudamiento de la municipalidad.
- Proyectos actuales que está llevando a cabo la municipalidad.
- Ubicación geográfica de puntos de interés (centros educativos, centros de salud y otros).

A continuación se muestra una funcionalidad del portal de datos abiertos de la Municipalidad de Pérez Zeledón que permite el uso de elementos de ubicación geográfica (a través de Google Maps).

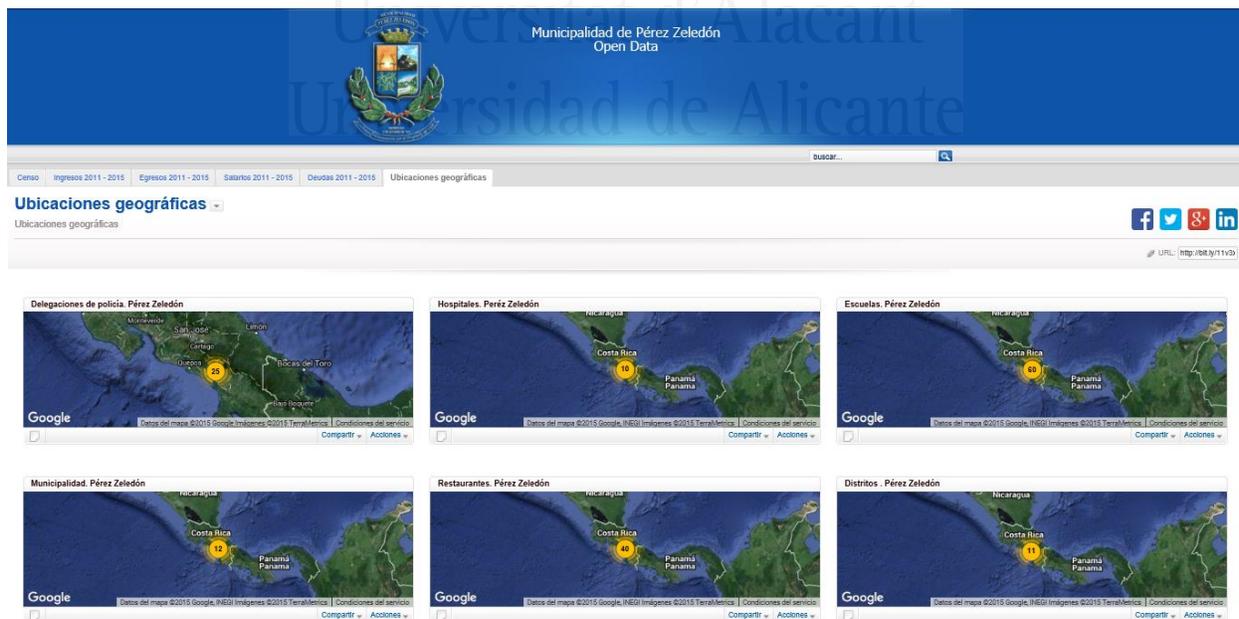


Figura 16: Ubicaciones geográficas en el portal de datos abiertos de la Municipalidad de Pérez Zeledón

⁴³ <http://opendata.mpz.go.cr>

Tras la aplicación del modelo de madurez al portal de datos abiertos de la Municipalidad de Pérez Zeledón se concluye que dicho portal alcanza el **Nivel 1 –G/I**, es decir, Básico sin cumplir Granularidad e Integración. A continuación los resultados obtenidos por cada criterio.

Criterio de Calidad	Resumen	¿Cumple Nivel 1?
D/R Disponibilidad y Reutilización	Se muestran conjuntos de datos disponibles y exportables independientemente	Sí
REL Relevancia	Se muestra información suficientemente representativa de su entorno	Sí
REP Reputación	Se indican las fuentes de datos y se respeta su contenido original	Sí
G/I Granularidad e Integración	Los datos se muestran con el máximo nivel de detalle posible	Parcialmente
V/U Visualización y Usabilidad	Se cuenta con herramientas para analizar la información usando gráficas	Sí

Tabla 15: Resultados de la aplicación del Modelo de Madurez al Portal de la Municipalidad de Pérez Zeledón

A continuación un breve análisis de la evaluación aplicada:

- En cuanto a **Visualización y Usabilidad**, el portal mostró características suficientes para cumplir con el requisito correspondiente al nivel 1, pero no fue posible observar que permita visualizar los mismos conjuntos de datos en diferentes formatos, lo cual le impide cumplir con el requisito de ese criterio para aspirar al nivel siguiente.
- En cuanto a **Granularidad e Integración**, tal y como se indicó sólo se cumple de manera parcial el requisito del nivel 1, pues no se evidenció que cuente con el mayor nivel de detalle posible en varios de los conjuntos de datos.
- En cuanto a la **Reputación**, el portal cumple tanto con el requisito para el nivel 1 como para el nivel 2, pues sus fuentes se indican y además son fuentes oficiales con respecto a los temas que abarca el portal.
- En cuanto a la **Relevancia**, solamente fue posible verificar el cumplimiento del requisito del nivel 1, pues aunque el portal muestra información representativa del entorno municipal, éste aún no alcanza la suficiente frecuencia de consulta como para ser considerado una fuente referente en su ámbito (requisito para el nivel 2).
- En cuanto a **Disponibilidad y Reutilización**, el portal cumple con el requisito de permitir exportar información de manera independiente para cada conjunto de datos (lo que lo ubica en nivel 1), pero no se evidenció que permita exportar información combinada de diferentes conjuntos de datos, por lo que no cumple con el requisito de ese criterio para alcanzar el siguiente nivel.

5.2 Portal de Datos Abiertos de la Municipalidad de Palmares

Como segundo caso de estudio se evaluó el portal de datos abiertos de la Municipalidad de Palmares⁴⁴, un pequeño cantón de la provincia de Alajuela que está ubicado en el occidente del Área Metropolitana de Costa Rica y cuenta con una población cercana a los 37.500 habitantes.

Al igual que el caso anterior, este portal de datos abiertos está implementado sobre la plataforma JUNAR lo que hace que existan muchas similitudes en cuanto al formato en que se publican los datos. No obstante, hay algunas diferencias importantes que radican principalmente en que:

- El contenido de este portal es mucho más completo.
- Los datos de este portal están presentados de forma mucho más detallada.
- Este portal ofrece una API para que los desarrolladores puedan disponer de los datos de manera automática (sin intervención manual).

A continuación se muestra la página inicial del portal de datos abiertos de la Municipalidad de Palmares.



Figura 17: Portal de datos abiertos de la Municipalidad de Palmares, Costa Rica

⁴⁴ <http://datos.munipalmares.go.cr>

Tras la aplicación del modelo de madurez al portal de datos abiertos de la Municipalidad de Palmares se concluye que dicho portal alcanza el **Nivel 1**, es decir, Básico. A continuación los resultados obtenidos por cada criterio.

Criterio de Calidad	Resumen	¿Cumple Nivel 1?
D/R Disponibilidad y Reutilización	Se muestran conjuntos de datos disponibles y exportables independientemente	Sí
REL Relevancia	Se muestra información suficientemente representativa de su entorno	Sí
REP Reputación	Se indican las fuentes de datos y se respeta su contenido original	Sí
G/I Granularidad e Integración	Los datos se muestran con el máximo nivel de detalle posible	Sí
V/U Visualización y Usabilidad	Se cuenta con herramientas para analizar la información usando gráficas	Sí

Tabla 16: Resultados de la aplicación del Modelo de Madurez al Portal de la Municipalidad de Palmares

Este portal de datos abiertos está implementado sobre la misma plataforma que el del caso anterior y además se refiere al mismo entorno (municipal), lo que origina que el cumplimiento de varios de los requisitos que establece este modelo de madurez tenga un comportamiento muy similar en ambos casos. No obstante, hay algunas diferencias importantes que radican principalmente en que el contenido del portal de la Municipalidad de Palmares, pues es mucho más completo y está presentado de forma mucho más granular. Justamente esa granularidad es la que propicia que en este portal el requisito “G/I Granularidad e Integración” sí se cumpla, lo cual le permite ubicarse en el Nivel 1 o Básico.

5.3 Portal de Datos Abiertos del MICITT

El portal de datos abiertos del Ministerio de Ciencia, Tecnología y Telecomunicaciones⁴⁵ (MICITT) fue desarrollado e implementado en el año 2014, en la actual administración. Es considerado uno de los portales de datos abiertos más completos y actualizados del país, pues dentro de su ámbito de acción, es decir la ciencia, la tecnología y las telecomunicaciones, abarca prácticamente todos los sectores, dedicando inclusive un sector del portal a cada uno de ellos:

- Actividades científicas y tecnológicas
- Investigación y desarrollo
- Recursos humanos en actividades de investigación y desarrollo
- Tecnologías de información y comunicación
- Innovación

⁴⁵ <http://indicadores.micit.go.cr>

A continuación se presenta una imagen que muestra la página principal del portal de datos abiertos del Ministerio de Ciencia, Tecnología y Telecomunicaciones de Costa Rica (MICITT).



Figura 18: Portal de datos abiertos del MICITT

La cantidad de conjuntos de datos que muestra este portal de datos abiertos es bastante amplia, y hace uso de elementos visuales que lo hacen bastante sencillo de interpretar.

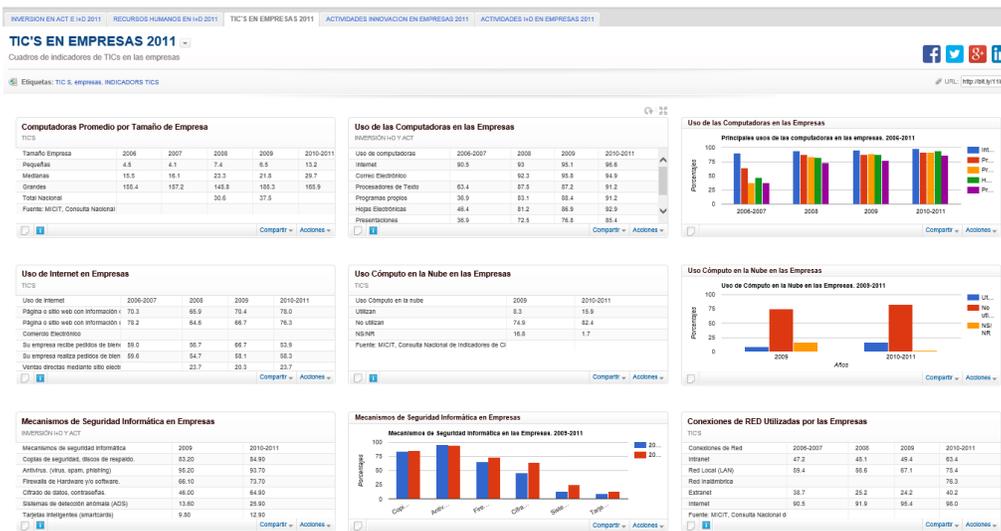


Figura 19: Conjuntos de datos sobre el tema TIC en el portal de datos abiertos del MICITT

Luego de haber aplicado el modelo de madurez propuesto al portal de datos abiertos del Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT), también implementado en JUNAR, se determinó que alcanza el Nivel 1 (Básico). Sin embargo, este portal presenta la particularidad de que cumple con varios requisitos de los siguientes niveles de madurez, por lo cual a continuación se muestra el análisis de los resultados detallando el nivel que logró alcanzar en cada uno de los criterios de calidad del modelo.

Criterio de Calidad	Nivel Alcanzado	Análisis
D/R Disponibilidad y Reutilización	1	El portal cumple con el nivel básico por presentar los datos en un formato abierto y permitir su análisis y exportación de forma independiente. No alcanza un nivel superior dado que no permite integrar distintos conjuntos de datos ni exportarlos de forma combinada.
REL Relevancia	3	El portal cumple con el nivel básico al considerarse que el volumen y la amplitud de los datos publicados en éste reflejan de manera muy representativa todos los ámbitos de acción en los que está involucrado este ministerio. Además, según pudo constatar, este portal de datos abiertos es altamente referenciado por otros sitios relacionados con ciencia, tecnología y telecomunicaciones, y de hecho se considera el sitio oficial de consulta de indicadores sobre estos temas a nivel nacional, por lo cual en el criterio de Relevancia puede considerarse que cumple también con los niveles 2 y 3 de madurez.
REP Reputación	2	Las fuentes de datos que alimentan a este portal de datos abiertos son indicadas debidamente y se respeta su contenido original, por lo cual se considera que cumple con el nivel 1 de madurez. Dichas fuentes además se pueden considerar como las fuentes oficiales de los temas que se publican dentro del portal, por lo cual también se cumple con el nivel 2 de madurez. Sin embargo, la frecuencia de actualización no está predefinida (o al menos no se indica), por lo cual no se puede otorgar el nivel 3 de madurez a este portal de datos abiertos, al no poderse asegurar que quien acceda y analice datos siempre contará con información vigente al momento en que realice la consulta.
G/I Granularidad e Integración	1	Al analizar el nivel de detalle con el que permite acceder y analizar la información este portal de datos abiertos, se puede concluir que cumple con el nivel 1 de madurez. Sin embargo, cuando se realiza una consulta específica no se pudo constatar que el portal refiera a otros conjuntos de datos relacionados ni tampoco que permita incorporar datos externos para incluirlos en el análisis, por lo cual no se cumple con los niveles 2 ni 3 del modelo de madurez.
V/U Visualización y Usabilidad	1	Se pudo determinar que el portal cumple con el nivel 1 de madurez dado que cuenta con suficientes elementos visuales de apoyo a los procesos de análisis por parte de los usuarios. No obstante, no fue posible constatar que la misma información pueda ser visualizada en distintos formatos, lo que le impide cumplir con el siguiente nivel de madurez. Tampoco se corroboró la existencia de herramientas o funcionalidades que permitan realizar procesos avanzados de análisis de datos, por lo que aun si hubiese cumplido con el nivel 2 (lo cual no sucede), no cumpliría con el requisito del nivel 3 de madurez.

Tabla 17: Resultados de la aplicación del Modelo de Madurez al Portal del MICITT

Se puede observar que el portal de datos abiertos del MICITT no alcanza el nivel 2 debido exclusivamente a limitaciones en cuanto a las funcionalidades del portal y no por la calidad de los datos en él publicados.



5.4 Conclusiones de la aplicación del modelo de madurez

Producto de la aplicación del modelo de madurez a los portales de datos abiertos de dos municipios y de un ministerio, todos pertenecientes a organizaciones costarricenses, se puede concluir lo siguiente:

- a. A pesar que los tres portales de datos abiertos están implementados en la misma plataforma (JUNAR), los resultados de la aplicación del modelo fueron distintos, sobre todo debido a los criterios de Relevancia, de Reputación y de Granularidad, los cuales trascienden el tema de la herramienta en la que estén desarrollados y tienen relación más directamente con la calidad del contenido publicado en estos.
- b. Particularmente con respecto al criterio de granularidad, pudo notarse cómo entre dos portales correspondientes a instituciones del mismo tipo (municipios) fue justamente ese criterio el que permitió a una municipalidad (la de Palmares) lograr un nivel superior que la otra (Pérez Zeledón, al permitir acceder con mayor detalle los datos publicados).
- c. Los portales de datos abiertos de los dos municipios únicamente aprobaron los requisitos correspondientes al nivel 1 (incluso hubo uno de ellos que no cumplió uno de los requisitos), mientras que el portal del MICITT alcanzó niveles superiores en dos temas específicamente: Relevancia y Reputación. Esto se debe a la calidad del contenido publicado, específicamente a que este portal de datos abiertos es considerado un referente en su campo y a que las fuentes de las que se alimenta son de alta confiabilidad.
- d. Se evidencia que al no permitir la herramienta en la que están implementados estos portales la integración ni la exportación conjunta de conjuntos de datos, existe una limitante considerable para incrementar el nivel en cuanto al criterio de Disponibilidad y Reutilización. Algo similar sucede con el criterio de Visualización y Usabilidad, pues los portales de datos abiertos evaluados no cuentan con la capacidad de visualizar los mismos conjuntos de datos en diferentes formatos ni con herramientas para realizar labores de análisis avanzado de la información.



Universitat d'Alacant
Universidad de Alicante



Capítulo 6: Incorporación del modelo de madurez en la Norma Técnica Nacional INTE 27-02-02:2015 para Portales de Datos Abiertos

Se muestran en este capítulo detalles generales sobre la Norma Técnica Nacional para Portales de Datos Abiertos para Costa Rica, incluyendo la incorporación del modelo de madurez desarrollado en esta tesis, así como los resultados de su aplicación al Portal de Datos del Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT) de Costa Rica.



Universitat d'Alacant
Universidad de Alicante

En diciembre de 2014, estando en curso el presente trabajo de investigación, el Instituto de Normas Técnicas de Costa Rica (INTECO) aprobó el desarrollo de una norma técnica que permitiera valorar y evaluar los portales de datos abiertos que, cada vez con más frecuencia, las instituciones públicas y algunas organizaciones de índole privado han venido implementando en este país, gracias al impulso que el Gobierno ha venido dando a este tema.

Para llevar a cabo la tarea de desarrollar dicha norma se creó un subcomité el cual fue presidido desde el inicio por el autor de esta tesis doctoral (Edgar Oviedo Blanco), y del cual formaron parte otros profesionales costarricenses relevantes relacionados con diversas áreas de la informática, quienes representaron a organizaciones tanto del sector público como privado.

6.1 Norma Técnica INTE 27-02-02:2015

La primera versión de la norma INTE 27-02-02:2015 fue finalizada en diciembre de 2015, coincidiendo con la finalización de la redacción de esta tesis doctoral, lo cual permitió que la misma pudiera ser aplicada en modo piloto a un portal de datos abiertos y sus resultados fueran analizados como parte de esta memoria de investigación.

En cuanto a su estructura, la norma inicia con una introducción que brinda un contexto general de la necesidad y la aplicabilidad de la norma dada la realidad actual del tema de datos abiertos en Costa Rica, así como los antecedentes que justifican el desarrollo de este valioso instrumento. Además, contempla los siguientes aspectos:

1. Objeto y campo de aplicación

Se especifica que la norma aplicará para portales de datos abiertos tanto de organizaciones de carácter público como privado.

2. Normas de referencia

Se hace referencia a algunas normas y documentos que fueron considerados como parte del desarrollo de esta norma técnica.

3. Términos y definiciones

En este apartado se definen más de treinta conceptos que son claves para el pleno entendimiento de la norma. Entre ellos se pueden encontrar “conjunto de datos”, “diccionario de datos”, “metadato”, “URL” y otros.

4. Pre-requisitos

La norma establece algunos pre-requisitos, los cuales en esencia consisten en una lista de condiciones operativas básicas que el portal deberá cumplir para satisfacer con la norma. Se incluyen aspectos como la alta disponibilidad, contar con un responsable de la publicación y gestión de contenidos del portal, y asegurar la integridad de los datos, entre otros.

5. Requisitos (preparación, formatos y contenido)



Este apartado es el que resume los principales temas que se evalúan, y de hecho, es importante mencionar que todos los requisitos de nivel 1 del modelo de madurez surgido de esta investigación doctoral están abarcados en la norma, la mayoría en este apartado, el cual clasifica los requisitos en varias categorías:

- a) Preparación de los datos
- b) Formato de los datos
- c) Compresión y empaquetamiento
- d) Contenido

6. Publicación

Se indican cuáles son los requisitos asociados específicamente con la publicación de los datos en el portal, contemplando aspectos como el uso de metadatos, el uso de licencias, la estructura del portal, el catálogo y otros detalles.

7. Evaluación de la Madurez del Portal de Datos Abiertos

Es el modelo de madurez que se propone como producto de esta investigación. Puede ser aplicado de manera complementaria a aquellos portales de datos abiertos que, más allá de cumplir con lo establecido en la prosa de la norma (y por ende alcanzar el nivel 1), deseen aspirar a tener un nivel de madurez mayor (2 o 3).

8. Correspondencia

Se hace mención de cuáles normas nacionales o internacionales tienen relación directa con esta norma técnica.

6.2 Incorporación de los resultados de la tesis a la norma

La norma INTE 27-02-02:2015 incluye una serie de requisitos – escritos de manera narrativa – los cuales deben ser aprobados de conformidad a lo establecido para cumplir con ésta. Dentro del cuerpo de la norma se consideraron los cinco requisitos que establece el modelo de madurez como parte de su nivel 1 o Básico. Es decir, si al evaluar un portal de datos abiertos con la norma INTE 27-02-02:2015 éste cumple de conformidad con todos los requisitos, se puede decir que cumple también con el nivel 1 o Básico del modelo de madurez.

Además de esto, se creó un apartado específico dentro de la norma (capítulo 7) para el modelo de madurez, que es el que permite a los portales de datos abiertos ser sometidos a la verificación del cumplimiento de los requisitos de los niveles 2 y 3, con el fin justamente de alcanzar dichos niveles de madurez.

Esta norma técnica estará disponible para ser utilizada oficialmente a inicios del año 2016. No obstante, como parte del trabajo coordinado y complementario que se ha venido realizando para la incorporación de los resultados de investigación de esta tesis doctoral a la norma técnica, se realizó una aplicación piloto de la norma y a continuación se muestran los resultados experimentales de dicha aplicación.

6.3 Resultados de la aplicación de la norma

A continuación se presentan, de forma detallada, los resultados del proceso de aplicación de la norma técnica INTE 27-02-02:2015 al Portal de Datos Abiertos del Ministerio de Ciencia, Tecnología y Telecomunicaciones (MICITT). Para facilitar la aplicación del proceso de evaluación de conformidad y la documentación de los resultados, todos los requisitos considerados de cumplimiento obligatorio fueron extraídos del cuerpo de la norma y ubicados en un formato de lista de chequeo.

1. Toda organización interesada en publicar conjuntos de datos de forma abierta debe contar con un portal dedicado a ello, o bien una sección de su portal organizacional donde se garantice las condiciones para la continuidad de los servicios.

Cumple No cumple

Comentarios: El MICITT cuenta con un portal de datos abiertos, el cual se puede acceder directamente en <http://indicadores.micitt.go.cr>. Se hace la observación de que el portal no es accesible desde el sitio oficial del MICITT⁴⁶, donde sí existen enlaces para acceder a otros portales, tal y como se muestra a continuación:



2. El portal debe estar operativo 24x7, es decir, de forma ininterrumpida.

Cumple No cumple

Comentario: Se verificó que el acceso al portal no tiene restricciones de tiempo, realizando ingresos en cinco diferentes momentos del día: 6:30 am, 10 am, 12:45 pm, 3:30 pm y 9:30 pm.

3. El portal debe poder ofrecer sus servicios a múltiples usuarios simultáneos.

Cumple No cumple

Comentario: Para verificar esto se abrieron con éxito diferentes sesiones en varias ventanas de navegadores e incluso en diferentes computadoras.

4. El portal debe contar con procedimientos de operación que permitan asegurar un nivel mínimo aceptable de calidad del portal.

Cumple No cumple

⁴⁶ www.micitt.go.cr

Comentario: Se verificó que el MICITT cuenta con procedimientos de operación específicos para la gestión del portal de datos abiertos.

5. El portal debe tener un responsable de gestionar el portal.

Cumple **No cumple**

Comentario: Se pudo constatar que existe una persona asignada como responsable de gestionar el portal de datos abiertos del MICITT.

6. El portal debe controlar el acceso para que sólo las personas o sistemas autorizados puedan publicar y modificar los datos y metadatos.

Cumple **No cumple**

Comentario: Se pudo constatar que solamente el responsable de gestionar el portal de datos abiertos puede publicar y/o modificar los datos y los metadatos.

7. El portal debe asegurar la integridad de los datos para que no sean alterados durante la transferencia.

Cumple **No cumple**

Comentario: Sí se pudo verificar que la exportación de los datos se realiza de manera completa, sin embargo, hubo una alteración parcial en el caso de caracteres especiales cuando se exportó a formato CSV, tal y como se muestra a continuación:

Zona Geográfica	Total	Hombres	Mujeres
Total 2011 (1)	571	393	164
América Latina	36	25	11
Costa Rica	122	71	51
USA y Canadá	162	122	40
Europa	219	158	61
Asia	1	1	0
Australia	1	1	0
Sin especificar (1)	30	15	1

Zona Geográfica	Total	Hombres	Mujeres
Total 2011 (1)	571	393	164
América Latina	36	25	11
Costa Rica	122	71	51
USA y Canadá	162	122	40
Europa	219	158	61
Asia	1	1	0
Australia	1	1	0
Sin especificar (1)	30	15	1

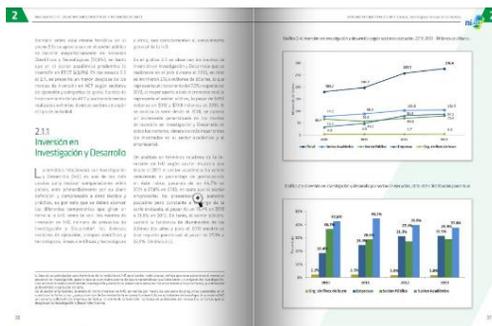
Datos Publicados

Datos Exportados (en XLS y en CSV)

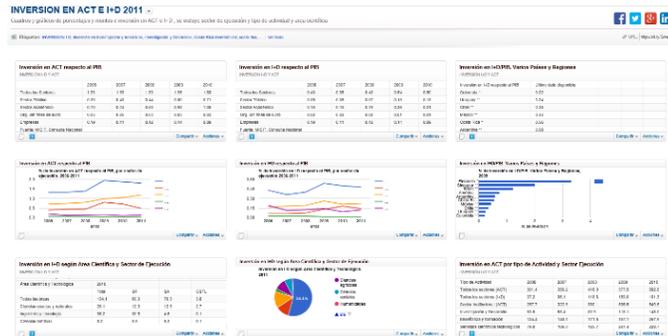
8. El portal debe asegurar que los datos provienen de la fuente de su publicador.

Cumple **No cumple**

Comentario: Se hizo un muestreo y se verificó que los datos publicados en los conjuntos de datos coinciden con los datos presentados en las fuentes originales indicadas (aunque en algunos casos se utilizan diferentes representaciones gráficas). A continuación un ejemplo de ello.



Fuente original indicada



Conjuntos de datos publicados

- El portal debe garantizar la permanencia en el tiempo de los datos que posee, de modo que todos los conjuntos de datos estén accesibles desde URLs que persistan en el tiempo y así evitar que se pierdan las referencias en el futuro.

X Cumple **___** No cumple

Comentario: Para cada conjunto de datos se establece un URL que permite acceder a éste de manera independiente.

- La alta dirección de la organización debe demostrar compromiso para garantizar la provisión de recursos necesarios para satisfacer los requerimientos técnicos del portal de datos abiertos.

X Cumple **___** No cumple

Comentarios: Se evidenció que existe presupuesto asignado por parte del MICITT para el soporte y mejoramiento continuo del portal de datos abiertos de la institución.

- Cada organización que impulse una iniciativa de datos abiertos debe incluir en su portal un mapa de navegación del portal, donde muestre la forma de acceder a todos sus conjuntos de datos.

X Cumple **___** No cumple

Comentario: Tal y como se muestra a continuación, en su página inicial el portal presenta un menú con las principales categorías en las que se clasifican los datos abiertos publicados, lo cual se considera suficiente para orientar al usuario en su afán de acceder a dichos datos.



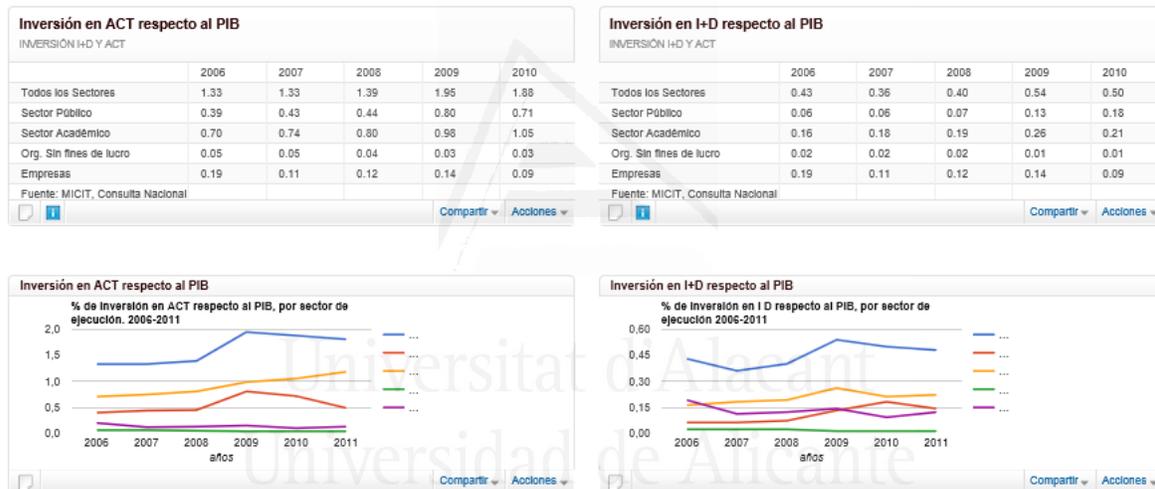
12. En caso que el portal sea administrado por un tercero, la organización dueña de la información debe asegurar que el tercero cuente con los requisitos establecidos para la administración del mismo (requisitos del 3 al 11).

No aplica.

13. Cada conjunto de datos debe estar publicado en la forma menos procesada posible, es decir, previos a cualquier agregación estadística o pre procesamiento. Además, la organización puede agregar recursos de datos en otras representaciones (visualizaciones, colecciones de datos, mapas) con base en estos datos.

Cumple **No cumple**

Comentario: Tal y como se muestra a continuación, la información se muestra en su estado más esencial y, además, se acompaña de otros elementos (en este caso gráficos) que complementan su análisis.



14. Cada conjunto de datos debe estar en formatos abiertos, no propietarios, estables y de amplio uso.

Cumple **No cumple**

Comentario: Se verificó que todos los datos abiertos publicados en el portal cumplen las características indicadas en este requisito.

15. Cada conjunto de datos debe tener un identificador único y permanente, según el estándar URI.

Cumple **No cumple**

Comentario: Se verificó la existencia de un identificador único para cada conjunto de datos.

16. Cada conjunto de datos debe especificar la información sobre sus datos y metadatos.

Cumple **No cumple**

Comentario: La información no es suficientemente específica en algunos de los conjuntos de datos publicados. Por ejemplo, en la siguiente imagen puede observarse que no se indica la unidad de medida en la que se presentan los datos (debería indicar que se trata de millones de dólares).

Inversión en I+D respecto al PIB

INVERSIÓN I+D Y ACT

	2006	2007	2008	2009	2010	2011
Todos los Sectores	0.43	0.36	0.40	0.54	0.50	0.48
Sector Público	0.06	0.06	0.07	0.13	0.18	0.14
Sector Académico	0.16	0.18	0.19	0.26	0.21	0.22
Org. Sin fines de lucro	0.02	0.02	0.02	0.01	0.01	0.01
Empresas	0.19	0.11	0.12	0.14	0.09	0.12
Fuente: MICIT, Consulta Nacional de Informe 2013: Cuadro 2.2						

17. En el caso de conjuntos de datos temporales, cada uno debe indicar su período de vigencia.

No aplica.

18. En el caso de conjuntos de datos con actualización periódica, cada uno debe indicar su periodicidad de actualización.

Cumple **No cumple**

Comentario: No se indica periodicidad de actualización en ningún conjunto de datos, y se nota que sí existen datos que requieren actualización periódica.

19. Cada conjunto de datos debe utilizar UTF-8 como mínimo para la codificación estándar de caracteres adoptada. Si se utiliza otro código se debe especificar el mismo en los metadatos.

Cumple **No cumple**

Comentario: El portal utiliza UTF-8 para la codificación de caracteres.

20. Cada conjunto de datos, para la denominación de los archivos, debe utilizar una convención de nombre estandarizado. Se recomienda el estilo "Upper Camel Case".

Cumple **No cumple**

Comentario: Se pudo notar que no existe una estandarización en cuanto al nombre de los archivos cuando se exportan los conjuntos de datos. De hecho, en algunos casos (no en todos) se utiliza archivos temporales de Excel con el prefijo "temp" pero cuando se utiliza la opción "Guardar como" el portal no sugiere ningún nombre para el archivo.

21. Si existe algún estándar con carácter oficial en el ámbito del portal con respecto a la estructura de los datos, cada conjunto de datos debe respetar todo convenio o esquema que esté previamente definido. Si se crean vocabularios o esquemas específicos de representación de la información, éstos se deben exponer en el portal para que se pueda interpretar correctamente la información al reutilizarla.

No aplica.

22. El portal debe proveer un canal eficiente de comunicación que permita la interacción bidireccional entre la organización y los consumidores de los datos.

Cumple No cumple

Comentario: Aunque el portal permite la publicación en las redes sociales Facebook, Twitter, Google+ y LinkedIn, realmente no provee en sí mismo un canal de comunicación interactiva con sus usuarios.

23. Las condiciones de uso del portal deben permitir la reutilización libre de los conjuntos de datos, incluso para fines comerciales. En aquellos casos donde exista una restricción de uso, se debe definir la licencia de acceso.

Cumple No cumple

Comentario: El portal cumple con este requisito al no establecer ninguna restricción en el consumo y reutilización de los datos publicados.

24. Los portales que contengan metadatos para coberturas y datos geo referenciados, deben tener asociados los metadatos indicados en el perfil oficial de metadatos geográficos de Costa Rica, emitido por el IGN basado en la norma ISO 19115.

No aplica.

25. Se debe tomar en cuenta lo prescrito por la Ley sobre Protección de la Persona frente al Tratamiento de sus Datos Personales.

Cumple No cumple

Comentario: No se observó ningún incumplimiento en este sentido.

26. En el caso de publicar datos genéricos, se deben utilizar formatos estándares y abiertos (no propietarios). En casos muy justificados (datos históricos, transformaciones costosas) es posible usar formatos propietarios o aún no estandarizados, pero indicando el software, versión y producto que permite procesar esos formatos.

Cumple No cumple

Comentario: Tal y como puede observarse a continuación, el portal muestra un menú emergente que permite exportar tanto en formatos abiertos (en CSV) como en formatos propietarios (Excel). En el segundo caso se indica que es la versión de Excel que se usa es 2007.



Investigadores según grado académico y sector de ejecución	
RECURSOS HUMANOS I+D	
Año 2011	3,970
Doctorado	571
Maestría y Especialidades	1,262
Licenciatura y Bachillerato	2,039
Otros grados académicos	98
Fuente: MICIT, Consulta Nacional	

Exportar a CSV
Exportar a Excel 2007
Código para incrustar
Obtener GUID
Google Spreadsheets

Compartir Acciones

27. El portal debe incluir información descriptiva y técnica (metadatos) sobre los conjuntos de datos que se exponen, que informen sobre cada conjunto de datos siguiendo una estructura común y estándar.

Cumple No cumple

Comentario: Se pudo constatar que en todos los conjuntos de datos se sigue un formato muy similar tanto para los datos como para los metadatos. Sin embargo, se notó algunas diferencias, tal y como se observa a continuación, donde en el primer caso sí se indica la URL de la fuente y en el otro no.

Inversión en I+D/PIB. Varios Países y Regiones

INVERSIÓN I+D Y ACT

Inversión en I+D respecto al PIB	Último dato disponible
Colombia **	0.22
Uruguay **	0.24
Chile **	0.35
México **	0.44
Costa Rica *	0.56
Argentina **	0.58
América Latina y el Caribe **	0.74
Brasil **	1.24
Singapur **	2.02
Finlandia *	3.32

Nota: * datos 2013, ** datos 2012
 Fuente: www.ricyt.org
<https://data.oecd.org/rd/gross-domes>
 Informe 2013: Gráfico 2.11

Inversión en I+D respecto al PIB + -

INVERSIÓN I+D Y ACT

	2006	2007	2008	2009	2010	2011
Todos los Sectores	0.43	0.36	0.40	0.54	0.50	0.48
Sector Público	0.06	0.06	0.07	0.13	0.18	0.14
Sector Académico	0.16	0.18	0.19	0.26	0.21	0.22
Org. Sin fines de lucro	0.02	0.02	0.02	0.01	0.01	0.01
Empresas	0.19	0.11	0.12	0.14	0.09	0.12

Fuente: MICIT, Consulta Nacional de
 Informe 2013: Cuadro 2.2

28. Los formatos de datos geográficos deben utilizar lo definido como oficial por el IGN para el sistema de referencia de coordenadas del país, actualmente denominado CRTM05, así como cualquier actualización o adecuación que se haga del sistema de referencia por parte del IGN. A nivel internacional la proyección válida es la denominada como WGS84.

No aplica.

29. Los formatos de datos geográficos deben ser compatibles con los estándares de interoperabilidad y normativa aplicada por el IGN. Se recomienda que la información sea compatible con el Sistema Nacional de Información Territorial (SNIT).

No aplica.

30. Para la publicación de la información geográfica tanto vectorial (orientados a los objetos) como raster (orientados a la superficie), se debe utilizar una de las siguientes opciones:

- Formatos estándares abiertos, cuando no se utilice una infraestructura de datos espaciales.
- Formatos que cumplen con los estándares definidos por la Open Geospatial Consortium (OGC), cuando se utilice una infraestructura de datos espaciales.

No aplica.

31. En casos donde la información se encuentre únicamente en formatos propietarios o aún no estandarizados, y no sea posible su transformación a formatos abiertos, se debe indicar el software, versión y productor que permite procesar esos formatos.

No aplica.

32. En caso de utilizar imágenes, se debe utilizar archivos basados en el método JPEG de compresión con pérdida (usualmente con extensión “.jpg”) y archivos basados en el método PNG (Portable Network Graphics) de compresión sin pérdida (usualmente con extensión “.png”).

No aplica.

33. En caso que se utilice formatos vectoriales y de visualización en la Web que necesiten representación, se debe utilizar el formato SVG (scalable vector graphics).

No aplica.

34. Para el caso específico de bases de datos estadísticos que no están en formato abierto, se debe publicar el archivo indicando el software, versión y productor que permite procesar esos formatos.

No aplica.

35. Cuando se utilice la compresión o el empaquetamiento de archivos, se debe hacer mediante una de las siguientes formas:

- Mediante compresión de archivos: En caso de ser necesario comprimir archivos de datos se debe usar algoritmos de compresión sin pérdida y abiertos (no propietarios), por ejemplo gzip, bzip2, ppmdi, entre otros.
- Mediante empaquetamiento de archivos: En caso de ser necesario empaquetar archivos de datos se debe utilizar algoritmos abiertos y estándares, por ejemplo tar. Adicionalmente se debe indicar explícitamente en los metadatos el protocolo utilizado.

No aplica.

36. El portal debe contener, como mínimo, los conjuntos de datos de mayor interés según las competencias de la organización. Los elementos principales que conforman la misión de la organización deben verse reflejados en la totalidad de los conjuntos de datos abiertos publicados en el portal.

Cumple **No cumple**

Comentario: Se pudo verificar que las diferentes secciones o páginas del portal, así como los conjuntos de datos publicados en cada una de ellas, representan conjuntamente la razón de ser del MICITT, al incorporar datos abiertos relacionados con Actividades Científicas y Tecnológicas, Investigación y Desarrollo, Recursos Humanos en Actividades de Investigación y Desarrollo, Tecnologías de Información y Comunicación e Innovación.



37. Los datos se debe mostrar con el máximo nivel de granularidad o detalle posible según el tema y el contexto de los datos publicados, considerando aspectos geográficos, demográficos y otros.

Cumple **No cumple**

Comentario: Se analizó cada conjunto de datos y se pudo verificar que todos muestran información a un nivel de detalle conveniente dentro de su contexto. Como ejemplos se pueden mencionar “Investigadores según grado académico y sector de ejecución” e “Investigadores por área científica y nivel académico”.

38. El portal debe contar con métricas internas que permitan evaluar sus indicadores de uso de los datos abiertos por parte de quienes los consuman.

Cumple **No cumple**

Comentario: No se pudo constatar la existencia de métricas internas de accesos y navegación dentro del portal.

39. El lenguaje que se debe utilizar para los datos y metadatos es Español-Costa Rica (ES_CR). Si además se dispone de información en otro idioma que es de interés para la organización, se recomienda publicarla.

Cumple **No cumple**

Comentario: Se pudo verificar que en todos los conjuntos de datos se emplea como idioma el español de Costa Rica.

40. La organización debe establecer las políticas o lineamientos generales en cuanto a cómo se va administrar, actualizar y publicar los datos abiertos así como también el tratamiento de los mismos.

Cumple **No cumple**

Comentario: Se verificó que el MICITT tiene establecidos procedimientos de operación específicos para la gestión del portal de datos abiertos, que incluyen los procesos indicados en este requisito.

41. Los metadatos obligatorios que debe incluir la publicación de un conjunto de datos son:

- a. Título: nombre definido para el conjunto de datos.
- b. Descripción: resumen descriptivo del conjunto de datos, especificando el contenido del recurso.
- c. Organización: nombre de la organización o entidad responsable de hacer el conjunto de datos públicamente disponible.
- d. Licencia: información sobre los derechos de uso del conjunto de datos.
- e. Fecha de publicación: punto de tiempo asociado al evento de publicación del conjunto de datos.
- f. Enlace de distribución: URL del archivo.

Cumple **No cumple**

Comentario: Los metadatos asociados a cada conjunto de datos no contemplan todo lo requerido. Tal y como se puede observar en la siguiente imagen, se pudo constatar que se incluye Título, Descripción, Organización (indicada como “Fuente”) y Fecha de Publicación (indicada como “Última

actualización”), sin embargo no se observa Licencia ni Enlace de distribución (URL del archivo de cada conjunto de datos).

Inversión Promedio en I+D por Tamaño de Empresa (\$)			
ACTIVIDADES I+D EMPRESAS			
Tamaño Empresa	2006	2007	2008
Pequeña	1,646	1,751	5,481
Mediana	7,419	5,399	20,699
Grande	76,608	46,447	74,108
Fuente: MICIT, Consulta Nacional			

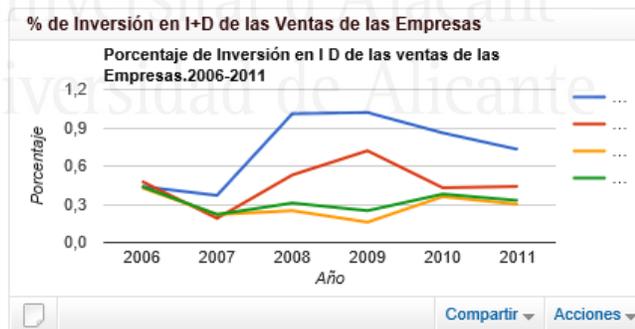
Última actualización: 10/08/2015 03:59

42. Cada uno de los recursos de datos derivados de los conjuntos de datos debe contener los siguientes metadatos:

- a. Enlace de distribución: URL del archivo.
- b. Tamaño: volumen físico utilizado por el archivo, en unidades estándares.
- c. Formatos: modos o medios principales de representación de los datos, normalmente identificados en la extensión del nombre del archivo.
- d. Descripción: características adicionales necesarias de conocer para su procesamiento.
- e. Fecha de publicación de los recursos derivados: fecha de publicación de esta distribución del archivo.

 Cumple X No cumple

Comentario: Los metadatos asociados a cada recurso de datos no contemplan todo lo requerido. Se pudo constatar el uso de Descripción solamente; no se observa ninguno de los otro metadatos requeridos (incluyendo la fecha de publicación que no se visualiza).



43. Aquellos conjuntos de datos que estén constituidos por uno o más archivos deben ser posibles de reconstituir a partir de los metadatos que se describen en la distribución del conjunto de datos.

No aplica.

44. En aquellos casos en que los datos son producto del procesamiento de otros conjuntos de datos, se debe indicar la procedencia de los datos que lo conforman.

No aplica.



45. Se deben utilizar licencias que permitan el libre uso de la información publicada en el portal, es decir, que no tenga barreras para su reutilización y difusión.

Cumple No cumple

Comentario: No se indica en el portal nada acerca del uso de alguna licencia, ni en la página inicial ni en ninguna de las páginas o secciones específicas.

46. Todo portal de datos abiertos debe contener como mínimo los siguientes componentes:

- a. Un catálogo que permita visualizar el contenido general de los datos.
- b. Facilidades, herramientas y estructuras para explorar los diferentes elementos del catálogo, permitiendo el uso básico de los datos publicados.

Cumple No cumple

Comentario: El portal cuenta con catálogo tanto en la página inicial como en las páginas o secciones específicas, así como con las facilidades y herramientas para su exploración.

47. Los catálogos de los conjuntos de datos publicados deben contener un listado exhaustivo de los conjuntos de datos presentes en el repositorio de la organización (es decir, todo dato publicado en el repositorio debe poder ser accesible a través del catálogo).

Cumple No cumple

Comentario: Aunque es simple, el catálogo principal de este portal de datos abiertos permite la navegación hacia todos los conjuntos de datos.

48. Si existen catálogos secundarios o específicos, estos deben estar enlazados con el catálogo central.

Cumple No cumple

Comentario: Los catálogos secundarios están en función del catálogo de la página principal.

49. El portal de datos abiertos debe proveer al menos las funcionalidades de navegación y búsqueda de contenidos, permitiendo observarlos, analizarlos, imprimirlos y exportarlos, ya sea como datos o como referencia a los datos para ser utilizados desde otros sitios web.

Cumple No cumple

Comentario: El portal permite navegar de forma sencilla para acceder todos los conjuntos de datos.

6.4 Conclusiones de la aplicación de la norma

Luego de haber aplicado la norma técnica INTE 27-02-02:2015 al portal de datos abiertos del MICITT, se obtuvieron las siguientes conclusiones:

- a. La evaluación lanzó los siguientes resultados para los 49 requisitos que conforman la norma:
 - Cumple: 25 requisitos
 - No cumple: 10 requisitos
 - No aplica: 14 requisitos



- b. Se puede observar que, considerando solamente los 35 requisitos que sí aplican dado el tipo de contenido que está publicado en este portal de datos abiertos del MICITT, la conformidad con la norma aplicada es de un 71%, un porcentaje relativamente alto pero insuficiente.
- c. Los requisitos no conformes o que no se cumplen giran alrededor de temas propios de la herramienta y no tanto a la calidad de los datos publicados, debiéndose en muchos casos a información que no se muestra en los conjuntos de datos o a funcionalidades que no brinda la herramienta.
- d. A pesar de abarcar muchos y muy diversos aspectos propios de un portal de datos abiertos, la aplicación de esta norma resultó ser una tarea relativamente sencilla, pues salvo algunos casos específicos donde se requiere verificar la existencia de elementos externos al portal (como los procedimientos de gestión o el responsable del portal, por ejemplo) fue posible la obtención de los resultados sin requerir un esfuerzo de indagación mayor al que se puede realizar simplemente navegando entre los conjuntos de datos abiertos publicados en el portal.
- e. A pesar de no cumplir con esta norma técnica, en la sección 5.3 de este documento se documentó que este mismo portal sí cumple con el nivel 1 o Básico del modelo de madurez generado a partir de esta tesis doctoral. Esto evidencia que existe una diferencia entre ambos instrumentos, la cual radica en que la norma contempla aspectos de diversa índole adicionales a los incluidos en el nivel 1 o Básico del Modelo de Madurez para Portales de Datos Abiertos.
- f. Al analizar el grado de rigurosidad de estos requisitos que conforman la norma se puede comprobar que los niveles 2 y 3 del modelo de madurez de portales de datos abiertos generado a partir de esta tesis de doctorado representan un reto adicional para los portales que se sometan a dicha evaluación.



Capítulo 7: Conclusiones y trabajo futuro

En este capítulo se exponen las conclusiones de todo el trabajo de investigación llevado a cabo a lo largo de esta tesis doctoral, haciendo hincapié en la incorporación de los resultados de investigación en la norma técnica nacional de Costa Rica promovida por INTECO.

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



7.1 Conclusiones

Este trabajo de investigación propone un modelo de madurez para portales de datos abiertos con cuatro niveles (Oviedo et al, 2015). Para alcanzar cada uno de estos niveles, un portal de datos abiertos que está siendo evaluado debe satisfacer una serie de criterios de calidad de datos abiertos, todos relacionados con alguno de los criterios de madurez establecidos. Estos criterios de calidad forman parte del modelo de calidad propuesto en (Oviedo et al, 2013). De esta manera, a partir de este modelo de madurez es posible calificar o ubicar un portal de datos abiertos en un nivel de madurez 0, 1, 2 o 3.

De manera paralela al curso de esta investigación, gracias al impulso que ha dado el Gobierno de Costa Rica al tema de Gobierno Abierto se logró el desarrollo de una norma técnica para la publicación de datos abiertos para la República de Costa Rica, basada en algunas de las normas ya existentes (principalmente la de Chile, tomando algunos detalles de la de España) y considerando el modelo de madurez definido en esta investigación como elemento distintivo, labor que fue llevada a cabo por el Instituto de Normas Técnicas de Costa Rica (INTECO), por medio del Subcomité Nacional de Ingeniería de Software de ese país, que consistió en un equipo interdisciplinario de profesionales costarricenses relacionados con el tema de Datos Abiertos, el cual fue presidido durante este tiempo por el autor de esta tesis doctoral (Édgar Oviedo Blanco).

Gracias a la existencia de un modelo de madurez y a la formalización de una norma técnica nacional que permite evaluar y definir si un portal de datos abiertos cumple con los requisitos mínimos que se esperan de una publicación de este tipo, hoy es posible aspirar en Costa Rica a que el nivel de calidad y de madurez que reflejen los múltiples portales de datos abiertos que con frecuencia se implementan en este país, mantengan una homogeneidad en cuanto a su calidad y a su nivel de madurez.

7.2 Trabajo futuro

Como trabajos futuros derivados o complementarios de esta investigación, se plantean los siguientes:

- a. Considerando que la empresa Grupo Babel cuenta con profesionales con conocimiento y experiencia en el campo de acción relacionado a esta investigación (implementación y desarrollo de portales así como aplicación de modelos de madurez), llevar a cabo labores de experimentación acerca de los criterios propuestos en el modelo de madurez, teniendo como sujetos de estudio precisamente a estos profesionales.
- b. A partir de los resultados de las tareas de experimentación anteriormente indicadas y considerando el criterio experto de otros investigadores, realizar una evaluación objetiva del modelo de madurez para determinar la precisión con que se están evaluando los portales de datos abiertos que se someten a dicho modelo.
- c. Incorporar mejoras al modelo de madurez para obtener una nueva versión con los conocimientos adquiridos a partir de la experimentación llevada a cabo.
- d. Desarrollar los procedimientos necesarios para que sea posible efectuar cálculos cuantitativos de las medidas de calidad que sustentan el proceso de evaluación de la madurez de los portales de datos abiertos.



Universitat d'Alacant
Universidad de Alicante



Capítulo 8: Resultados de la investigación

En este capítulo se enumeran los productos obtenidos como resultado de investigación obtenidos a lo largo de este proceso de investigación.

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



Memoria de Investigación del DEA

- Oviedo, E. (2013). Memoria de Suficiencia Investigadora: Definición de medidas de calidad para datos abiertos: caso de estudio de datos públicos en la Municipalidad de Pérez Zeledón, en Costa Rica. Universidad de Alicante, España, 2013.

Este documento, presentado en Matanzas, Cuba, en enero de 2013 como parte del proceso de suficiencia investigadora para obtener el Diplomado de Estudios Avanzados (DEA) otorgado por la Universidad de Alicante, recopiló los resultados de la investigación llevada a cabo durante el año 2012, la cual implicó un amplio estudio del estado del arte en el tema de datos abiertos y otros conceptos relacionados. Permitió contar con una primera definición de medidas o criterios de calidad para portales de datos abiertos, que luego representaron la base del modelo de madurez propuesto en el actual documento. Dichos criterios, incorporados como parte de un modelo de calidad, fueron aplicados al portal de datos abiertos de la Municipalidad de Pérez Zeledón, el cual fue implementado justamente para dicho fin.

Artículos presentados en Conferencias y Congresos

- Oviedo, E., Mazón J.N., Zubcoff, J.J. (2013) Towards a data quality model for open data portals. CLEI, 2013.

En este artículo, basado en la Memoria de Suficiencia Investigadora descrita anteriormente, se presentó de manera más específica el modelo de calidad que había sido previamente definido, así como los resultados de su aplicación al portal de datos abiertos de la Municipalidad de Pérez Zeledón. Fue presentado en la Conferencia Latinoamericana en Informática del año 2013⁴⁷, llevada a cabo en la ciudad de Naiguatá, Venezuela.

- Oviedo, E., Mazón J.N., Zubcoff, J.J. (2015) Quality and maturity model for open data portals. CLEI, 2015.

En este artículo se presenta por primera vez el modelo de madurez para portales de datos abiertos, el cual está sustentado en los criterios que formaron parte del modelo de calidad planteado en el trabajo previo. Fue presentado en la Conferencia Latinoamericana en Informática del año 2015⁴⁸, realizada en la ciudad de Arequipa, Perú.

⁴⁷ <http://www.clei.org/clei2013/programa-sistemas-inovadores-de-datos/index.html>

⁴⁸ <http://eventos.spc.org.pe/clei2015/pdfs/144712.pdf>



Artículos presentados en Revistas

- Oviedo, E., Mazón J.N., Zubcoff, J.J. (2015). Considering open data quality for measuring maturity of open data portals. CLEI Electronic Journal, 2015.

Como parte del proceso de evaluación de los artículos presentados para la Conferencia Latinoamericana en Informática del año 2015, el artículo que había sido presentado por el autor de esta investigación fue seleccionado como uno de los mejores trabajos para ser publicado en la revista electrónica del Centro Latinoamericano de Estudios en Informática, llamada CLEI Electronic Journal⁴⁹.

Otras publicaciones relacionadas

- INTECO (2015). Ingeniería de software y sistemas: Portales de Datos Abiertos – Requisitos. Norma Técnica Nacional para Portales de Datos Abiertos (INTE 27-02-02:2015), 2015.

Universitat d'Alacant
Universidad de Alicante

⁴⁹ CLEIej, ISSN 0717- 5000. <http://www.clei.org/cleiej>



Capítulo 9: Anexos

En este capítulo se incorporan tres anexos que contienen:

- A) El documento completo del Proyecto de Norma INTE 27-02-02:2015. Se hace hincapié en que esta norma debe ser sometida a un procedimiento de aprobación por parte de las autoridades de INTECO antes de ser oficialmente publicada como norma técnica.
- B) La formulación cuantitativa que sustenta una forma de calcular el valor para cada una de las medidas de calidad establecidas en el modelo de calidad presentado en (Oviedo et al 2013).
- C) Carta emitida por la Ing. Alexandra Rodríguez, Directora de Normalización de INTECO, donde se reseña la participación de autor de esta tesis doctoral en el desarrollo de la norma INTE 27-02-02:2015.



Universitat d'Alacant
Universidad de Alicante



Anexo A: Proyecto de Norma INTE 27-02-02:2015

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



INTE CTN 27 SC 02

Fecha: 2015-12-09

PN INTE 27-02-02:2015

Proyecto de Norma – Primera Edición

Secretaría: INTECO



Universitat d'Alacant
Universidad de Alicante

**Ingeniería de software y sistemas – Portales de Datos
Abiertos - Requisitos**

CORRESPONDENCIA:

Introducción

En los últimos años el tema de datos abiertos (“open data”) ha venido adquiriendo mayor interés por parte de las autoridades y de la población en general, alrededor de todo el mundo. Se conoce como datos abiertos a aquellos datos que se publican en sitios web de acceso público con el fin de ser reutilizados y redistribuidos por aquellas personas u organizaciones que lo deseen, sin tener ningún tipo de restricciones para ello.

Este innovador pero a la vez sencillo concepto se ha popularizado gracias a la necesidad y el deseo que tienen las personas de poder acceder a información fidedigna, completa y de manera oportuna relacionada con temas de interés general, tales como el uso de presupuestos públicos, proyectos de desarrollo económico y social de los pueblos, información demográfica, entre otros.

Los datos abiertos están disponibles a través de portales de datos abiertos que permiten su análisis, para una toma de decisiones informada. La implementación de portales de datos abiertos propicia la inclusión de datos provenientes de muy diversas fuentes de información, desde fuentes muy formales y estructuradas hasta fuentes más informales con información organizada de manera muy básica.

En el caso de Costa Rica en el año 2015, el Ministerio de la Presidencia lanzó una importante iniciativa llamada “Plan de Acción de Gobierno Abierto”, que tiene como objetivo primordial una plena implementación del Gobierno Abierto en este país. Además, mediante el decreto ejecutivo N° 38994-MP-PLAN-MICITT se creó la Comisión Nacional de Gobierno Abierto, la cual promueve políticas, lineamientos, estrategias y una metodología de evaluación en materia de Gobierno Abierto, y coordina las acciones necesarias para cumplir los principios de transparencia, rendición de cuentas y participación ciudadana.

Dentro de este contexto surgió una iniciativa para desarrollar una norma técnica dirigida a proveer los lineamientos para la publicación de datos abiertos que cuenten con un acceso expedito y sin restricciones de uso, en formatos estructurados, que permitan su reutilización tanto por terceros como de manera automatizada, en el entendido que esto puede significar un importante estímulo a entidades que pueden agregar valor a dicha información y, eventualmente, potenciar actividades de emprendimiento asociadas a la divulgación y agregación de información relevante para el ciudadano.

En Costa Rica existen diversos portales de datos abiertos principalmente implementados por ministerios, instituciones autónomas y municipios, que están disponibles para que los ciudadanos se enteren con absoluta transparencia de aspectos socialmente sensibles tales como la inversión de fondos públicos, los proyectos en desarrollo y todo tipo de información propia del entorno al que se refiere la organización. La presente norma pretende colaborar al país en su esfuerzo por la implementación del Gobierno Abierto, estableciendo los lineamientos básicos que permitan estandarizar la medición del nivel de calidad de los portales de datos abiertos que se implementen, así como un modelo de madurez que permita a todo portal de datos abiertos ubicarse en un nivel específico de madurez, a partir del cumplimiento de los requisitos establecidos en dicho modelo.



Ingeniería de software y sistemas - Portales de Datos Abiertos – Requisitos

1. OBJETO Y CAMPO DE APLICACIÓN

Esta norma nacional establece los lineamientos que se deben cumplir para facilitar el acceso y uso de la información pública a través de los portales de datos abiertos, mediante el establecimiento de las características mínimas que deben cumplir los conjuntos de datos así como el proceso de su publicación, entendiendo por éste el proceso de hacer exposición pública constante y actualizada de conjuntos y recursos de datos por parte de una organización para el uso abierto de todos los ciudadanos, ya sea por medio de interfaces humanas o procedimientos automatizados. Incluye la publicación de los datos propiamente, así como sus correspondientes metadatos.

Esta norma aplica a cualquier organización costarricense, pública o privada, que publique datos a través de portales de datos abiertos.

2. NORMAS DE REFERENCIA

Los siguientes documentos, en su totalidad o en parte, están referenciados en este documento y son indispensables para su aplicación. Para las referencias con fecha sólo se aplica la edición citada. Para las referencias sin fecha se aplica la última edición del documento de referencia (incluyendo cualquier modificación).

- ISO / TC 211
- ISO 19115
- DCAT
- SCHEMA.ORG
- LINKED.OPEN.DATA
- LICENCIAS

3. TÉRMINOS Y DEFINICIONES

Para los propósitos de esta norma se emplean los términos y definiciones mencionados a continuación:

base de datos

Cualquier archivo, fichero, registro u otro conjunto estructurado de datos que sea objeto de tratamiento o procesamiento, sea de forma manual o automatizada, independientemente de la modalidad de su elaboración, organización o acceso.

base de datos estadísticos

Datos recolectados y publicados para fines estadísticos. Están en formatos particulares para reúso y explotación de aplicaciones especializadas.

catálogo de datos



Inventario ordenado de datos disponibles en distintos formatos, que permite al usuario acceder y localizar la información vía web.

conjunto de datos

Colección de datos vinculados entre sí y agrupados para su potencial reutilización.

consumidor de datos

Cualquier persona u organización que accede datos publicados y obtiene una copia total o parcial de ellos para fines propios.

dato

Representación de hechos, conceptos o instrucciones de un modo formalizado y adecuado para su comunicación, interpretación o procesamiento por medios automáticos o humanos.

datos abiertos (open data)

Datos que no tienen restricciones de acceso de ningún tipo, en particular ni administrativas ni tecnológicas.

datos enlazados (linked data)

Datos que están enlazados por protocolos informáticos, en particular por medio de URIs o enlaces href.

datos genéricos

Datos que no tienen aplicaciones especializadas y que normalmente se corresponden con datos provenientes de bases de datos y reportes, como planillas de cálculo y cuadros de información.

datos geográficos

Entidades espacio-temporales que cuantifican la distribución, el estado y los vínculos de los distintos fenómenos u objetos naturales y sociales.

datos personales

Cualquier dato relativo a una persona física identificable. Los elementos básicos son:

- i. Información relativa a una persona, siendo indiferente la naturaleza del dato, antecedente o hecho de que se trate.
- ii. Información que permita identificar al titular. Se entiende para estos efectos por identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, mediante un número de identificación, o mediante elementos específicos característicos de su identidad física, fisiológica, psíquica, económica, cultural o social, tales como identificación, número de cuenta bancaria, domicilio, número telefónico, entre otros.

Quedan comprendidos dentro de esta definición, independientemente del soporte en que se encuentren, datos tales como: nombre, edad, sexo, identificación, estado civil, profesión, domicilio, números telefónicos, imagen, entre otros.

datos primarios (raw data)

Expresión que refiere a datos en estado “original”, es decir, que no han sido sujeto de procesamiento, transformación, agregación o cualquier otra manipulación.

datos sensibles

Datos personales que se refieren a las características físicas o morales de las personas, o a hechos o circunstancias de su vida privada o íntima, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud, físicos o psíquicos y la vida sexual.

derecho de autor

Derecho moral, irrenunciable, inalienable y exclusivo de explotación que tiene y ejerce sobre su obra literaria, artística o científica, la persona creadora de la misma.

derecho de propiedad intelectual

Derecho de autor y derechos afines, incluidas las formas de protección sui géneris.

diccionario de datos

Descripción del esquema y taxonomías usadas en los conjuntos de datos.

disociación de los datos

Todo tratamiento de datos personales de manera que la información que se obtenga no pueda asociarse a una persona determinada o determinable.

disponibilidad

Propiedad o característica de los activos, consistentes en que las entidades o procesos autorizados tienen acceso a los mismos cuando lo requieren.

documento

Toda información, cualquiera que sea su soporte material o electrónico, así como su forma de expresión gráfica, sonora o en imagen utilizada.

esquema (de datos)

Definición de conjunto de campos de diferentes tipos que tienen por objetivo registrar y organizar metadatos.

formato

Conjunto de características técnicas y de presentación.

formato de datos

Estructura física y lógica usada para almacenar datos en un archivo. Usualmente se indica por un sufijo al final del nombre del archivo. Ejemplo: miArticulo.pdf, publicaciones.xml.

información

Hecho conocido, con significado explícito. La información consta de datos más metadatos que los describen.

mapa de navegación

Proporcionan una representación esquemática de la estructura del hipertexto, indicando los principales conceptos incluidos en el espacio de la información y las interrelaciones que existen entre ellos.

Nota: El mapa de navegación puede representarse en forma textual, en forma gráfica, o mediante una combinación de ambas. Un mapa es, por ejemplo, una representación completa (o resumida) del sitio web para orientar al lector/usuario durante el recorrido o para facilitarle un acceso directo al lugar que le interese.

metadatos

Datos utilizados para suministrar información sobre datos producidos. Describen el contenido, calidad, condiciones, historia, disponibilidad y cualquier aspecto que se considere necesario para ampliar la conceptualización del dato descrito

portales de datos abiertos

Sitios web que publican datos de acceso libre con el fin de ser reutilizados y redistribuidos sin restricciones por aquellas personas u organizaciones que lo deseen.

publicación de datos abiertos

Proceso de hacer exposición pública en Internet de los conjuntos de datos por parte de una organización, una vez preparados para el uso abierto de todos los usuarios. Incluye la publicación de los datos propiamente, así como sus correspondientes metadatos.

La publicación de datos abiertos es un modelo de intercambio de datos abiertos que no requiere acuerdos bilaterales previos entre el publicador del dato y el consumidor del dato. Los datos están disponibles bajo formatos y patrones de proceso conocidos y estándares. El objetivo es que cualquier contraparte pueda usar los datos y metadatos.

publicador

Organización u individuo dentro de la organización que se hace responsable por la publicación de datos.

recursos de datos derivados

Derivados de datos que pueden constituirse en tablas, visualizaciones, mapas, infografías, gráficos, colecciones temáticas de datos u otros, que se pudieran desarrollar con base en los datos primarios o conjuntos de datos.

repositorio (de datos)

Unidad virtual de los conjuntos de datos físicos publicados por una organización.

responsable de la base de datos

Persona física o jurídica que administra, gerencia o se encarga de la base de datos, ya sean de una entidad pública o privada, competente para decidir cuál es la finalidad de la base de datos, cuáles categorías de datos deberán registrarse o publicarse y qué tipo de tratamiento se les aplicará.

reutilizar



Volver a utilizar algo previamente existente, ya sea con la función que desempeñaba anteriormente o con otros fines.

schema.org

Esfuerzo colaborativo y comunitario con la misión de crear, mantener y promover esquemas adecuados para publicación de datos estructurados en internet, en páginas web, en email y más.

Nota. El vocabulario definido por Schema.org puede ser utilizado en muy diferentes tipos de codificación incluyendo RDFa, Microdata y JSON-LD. Estos vocabularios cubren entidades, relaciones entre entidades y acciones. Los mismos pueden ser extendidos siempre y cuando se documenten bien dichos modelos de extensión.

servicio público

Actividad llevada a cabo por la Administración o bajo un cierto control y regulación de ésta, por una organización, especializada o no, y destinada a satisfacer necesidades de la colectividad.

sistema

Conjunto de soluciones tecnológicas (informáticas) que permiten la organización, almacenamiento y recuperación de datos.

transparente

Algo que es claro, evidente; que se comprende sin duda ni ambigüedad.

Upper Camel Case

Convención de nombre en la cual muchas palabras se juntan en una sola, donde la primera letra de la primera palabra se escribe con mayúscula y las primeras letras de las siguientes palabras se escriben con mayúscula. Ejemplo: LluviasSeptiembre2010.xml

URI (Uniform Resource Identifier)

En español "identificador uniforme de recurso", es una cadena de caracteres corta que identifica inequívocamente un recurso (servicio, página, documento, dirección de correo electrónico, enciclopedia, etc.). Normalmente estos recursos son accesibles en una red o sistema. Los URI pueden ser localizadores uniformes de recursos (URL), nombres uniformes de recursos (URN), o ambos (URL + URN).

URL (Uniform Resource Locator)

Nombre o identificador. Secuencia compacta de caracteres que permite localizar un recurso describiendo su modo primario de acceso. Las URL son un subconjunto de las URI.



4. PRE-REQUISITOS

Toda organización interesada en publicar conjuntos de datos de forma abierta debe contar con un portal dedicado a ello, o bien una sección de su portal organizacional donde se garantice las condiciones para la continuidad de los servicios. Los elementos que se deben cumplir son los siguientes:

- a) estar operativo 24x7;
- b) ofrecer un servicio a múltiples usuarios simultáneos;
- c) contar con procedimientos de operación que garanticen el cumplimiento de las condiciones generales mencionadas en el capítulo 5 de esta norma, que permitan asegurar un nivel mínimo aceptable de calidad del portal;
- d) tener un responsable de gestionar el portal;
- e) controlar el acceso para que sólo las personas o sistemas autorizados puedan publicar y modificar los datos y metadatos;
- f) asegurar la integridad de los datos para que no sean alterados durante la transferencia;
- g) asegurar que los datos provienen de la fuente de su publicador;
- h) garantizar la permanencia en el tiempo de los datos que posee, de modo que todos los conjuntos de datos estén accesibles desde URLs que persistan en el tiempo y así evitar que se pierdan las referencias en el futuro.

La alta dirección de la organización debe demostrar compromiso para garantizar la provisión de recursos necesarios para satisfacer los requerimientos técnicos del portal de datos abiertos.

Cada organización que impulse una iniciativa de datos abiertos debe incluir en su portal un mapa de navegación del portal, donde muestre la forma de acceder a todos sus conjuntos de datos.

Adicionalmente, en caso que el portal sea administrado por un tercero, la organización dueña de la información debe asegurar que el tercero cuente con los requisitos establecidos anteriormente para la administración del mismo.

5. REQUISITOS (PREPARACIÓN, FORMATOS Y CONTENIDO)

Con el fin de facilitar el acceso expedito, abierto, y sin restricciones de uso de conjuntos de datos organizacionales, antes de la publicación el responsable organizacional debe asegurar que los conjuntos de datos cumplan con las siguientes condiciones generales:

5.1 Preparación

Cada uno de los conjuntos de datos debe cumplir con lo siguiente:

1. Estar publicado en la forma menos procesada posible, es decir, previo a cualquier agregación estadística o pre-procesamiento. Además de entregar la información sin pre-procesamiento, la organización puede agregar recursos de datos en otras representaciones (visualizaciones, colecciones de datos, mapas) con base en estos datos.
2. Estar en formatos abiertos, no propietarios, estables y de amplio uso.
3. Tener un identificador único y permanente, según el estándar URI.

4. Especificar la información sobre sus datos y metadatos (ver Capítulo 6).
5. Indicar el período de vigencia para conjuntos de datos temporales.
6. Indicar la periodicidad de actualización para conjuntos de datos con actualización periódica.
7. Utilizar UTF-8 como mínimo para la codificación estándar de caracteres adoptada. Si se utiliza otro código se debe especificar el mismo en los metadatos.
8. Para la denominación de los archivos se debe utilizar una convención de nombre estandarizado. Se recomienda el estilo *“Upper Camel Case”*.
9. Si existe algún estándar con carácter oficial en el ámbito del portal con respecto a la estructura de los datos, debe respetarse todo convenio o esquema que esté previamente definido. Si se crean vocabularios o esquemas específicos de representación de la información, éstos se deben exponer en el portal para que se pueda interpretar correctamente la información al reutilizarla.
10. El portal debe proveer un canal eficiente de comunicación que permita la interacción bidireccional entre la organización y los consumidores de los datos.
11. Las condiciones de uso del portal deben permitir la reutilización libre de los conjuntos de datos, incluso para fines comerciales; en aquellos casos donde exista una restricción de uso, se debe definir la licencia de acceso.
12. Los portales que contengan metadatos para coberturas y datos geo referenciados, deben tener asociados los metadatos indicados en el perfil oficial de metadatos geográficos de Costa Rica, emitido por el IGN basado en la norma ISO 19115.

Adicionalmente se debe tomar en cuenta lo prescrito por la Ley, sobre Protección de la Persona frente al Tratamiento de sus Datos Personales.

5.2 Formatos

Sobre los formatos de los conjuntos de datos, la presente norma agrupa cuatro grandes categorías, una genérica y tres especializadas, cuyas características técnicas exigen un tratamiento particular. Los formatos permitidos son los siguientes:

5.2.1 Datos genéricos

Se deben utilizar formatos estándares y abiertos (no propietarios).

Nota. En casos muy justificados (datos históricos, transformaciones costosas), es posible usar formatos propietarios o aún no estandarizados pero indicando el software, versión y producto que permite procesar esos formatos.

El portal debe incluir información descriptiva y técnica (metadatos) sobre los conjuntos de datos que se exponen, que informen sobre cada conjunto de datos siguiendo una estructura común y estándar.

5.2.2 Datos geográficos

Los formatos de datos geográficos deben cumplir con los siguientes requisitos:

1. Utilizar lo definido como oficial por el IGN para el sistema de referencia de coordenadas del país, actualmente denominado CRTM05, así como cualquier actualización o adecuación que se haga del sistema de referencia por parte del IGN. A nivel internacional la proyección válida es la denominada como WGS84.

2. Ser compatibles con los estándares de interoperabilidad y normativa aplicada por el IGN. Se recomienda que la información sea compatible con el Sistema Nacional de Información Territorial (SNIT).
3. Para la publicación de la información geográfica tanto vectorial (orientados a los objetos) como raster (orientados a la superficie), se debe utilizar una de las siguientes opciones:
 - a. Formatos estándar abiertos, cuando no se utilice una infraestructura de datos espaciales.
 - b. Formatos que cumplen con los estándares definidos por la Open Geospatial Consortium (OGC), cuando se utilice una infraestructura de datos espaciales.

Se recomienda adoptar la normativa establecida para la Infraestructura de Datos Espaciales de Costa Rica (IDECORI).

Adicionalmente, para la publicación de información geográfica se recomienda aplicar la normativa desarrollada por el comité internacional ISO / TC 211 (Geographic information/Geomatics). Y para facilitar el intercambio y la gestión de la información geográfica se recomienda utilizar las normas de información geográfica establecidas por el IGN.

En casos donde la información se encuentre únicamente en formatos propietarios o aún no estandarizados y no sea posible su transformación a formatos abiertos, se debe indicar el software, versión y productor que permite procesar esos formatos.

5.2.3 Imágenes

Si se utilizan imágenes para expandir o aclarar el conjunto de datos, se debe cumplir con lo siguiente:

- Utilizar formatos de imágenes basados en el método JPEG de compresión con pérdida, usualmente con extensión “.jpg”, y los archivos basados en el método PNG (Portable Network Graphics) de compresión sin pérdida, usualmente con extensión “.png”.
- Utilizar el formato SVG (scalable vector graphics) en caso en que se utilice formatos vectoriales y de visualización en la Web que necesiten representación.

Nota: Una imagen no es considerada como un dato abierto.

5.2.4 Base de datos estadísticos

Para el caso específico de bases de datos estadísticos que no están en formato abierto, se debe publicar el archivo indicando el software, versión y productor que permite procesar esos formatos.

5.3 Compresión y empaquetamiento de archivos

Cuando se utilice la compresión o el empaquetamiento de archivos, se debe hacer mediante una de las siguientes formas:

a) Mediante compresión de archivos:

En caso de ser necesario comprimir archivos de datos se debe usar algoritmos de compresión sin pérdida y abiertos (no propietarios), por ejemplo gzip, bzip2, ppmdi, entre otros.

b) Mediante empaquetamiento de archivos:



En caso de ser necesario empaquetar archivos de datos se debe utilizar algoritmos abiertos y estándares, por ejemplo tar. Adicionalmente se debe indicar explícitamente en los metadatos el protocolo utilizado.

Nota: Tener en cuenta que al comprimir o empaquetar datos, se genera un mayor esfuerzo para el usuario.

5.4 Contenido

El portal debe contener, como mínimo, los conjuntos de datos de mayor interés según las competencias de la organización. Los elementos principales que conforman la misión de la organización deben verse reflejados en los conjuntos de datos abiertos publicados en el portal.

Los datos se deben mostrar con el máximo nivel de granularidad o detalle posible según el tema y el contexto de los datos publicados, considerando aspectos geográficos, demográficos y otros.

El portal debe contar con métricas internas que permitan evaluar sus indicadores de uso de los datos abiertos por parte de quienes los consuman.

6. PUBLICACIÓN

El lenguaje que se debe utilizar para los datos y metadatos es Español-Costa Rica (ES_CR).

Nota. Si además se dispone de información en otro idioma se recomienda publicarla si es de interés para la organización.

La organización debe establecer las políticas o lineamientos generales en cuanto a cómo se va administrar, actualizar y publicar los datos abiertos así como también el tratamiento de los mismos.

6.1 Metadatos

6.1.1 Metadatos obligatorios

Los metadatos obligatorios que debe incluir la publicación de un conjunto de datos son:

- a. *Título:* nombre definido para el conjunto de datos.
- b. *Descripción:* resumen descriptivo del conjunto de datos, especificando el contenido del recurso.
- c. *Organización:* nombre de la organización o entidad responsable de hacer el conjunto de datos públicamente disponible.
- d. *Licencia:* información sobre los derechos de uso del conjunto de datos.
Nota. Para mayor información referirse al apartado referente a Licencias (apartado 6.2).
- e. *Fecha de publicación:* punto de tiempo asociado al evento de publicación del conjunto de datos.
- f. *Enlace de distribución:* URL del archivo.



A su vez, cada uno de los *recursos de datos derivados* de los conjuntos de datos debe contener los siguientes metadatos:

- a. *Enlace de distribución*: URL del archivo.
- b. *Tamaño*: volumen físico utilizado por el archivo, en unidades estándares.
- c. *Formatos*: modos o medios principales de representación de los datos, normalmente identificados en la extensión del nombre del archivo.
- d. *Descripción*: características adicionales necesarias de conocer para su procesamiento.
- e. *Fecha de publicación de los recursos derivados*: fecha de publicación de esta distribución del archivo.

6.1.2 Metadatos recomendados

Los metadatos recomendados para la publicación de un conjunto de datos son:

- a. *Etiquetas*: términos clave para describir el conjunto de datos. Pueden ser agregados de forma dinámica y de acuerdo a sus consideraciones.
- b. *Frecuencia de actualización*: frecuencia temporal con que se actualiza el conjunto de datos.
- c. *Plazo de vigencia*: periodo durante el cual el dato tiene vigencia.
- d. *Categoría*: clasificación conceptual básica del conjunto de datos en sistemas de categorías o taxonomías disponibles.
- e. *Cobertura geográfica*: zona(s) geográfica(s) asociada(s) al conjunto de datos, se debe especificar:
 - distrito,
 - cantón,
 - provincia,
 - país.
- f. *Cobertura temporal*: punto o periodo de tiempo al cual está asociado el contenido del conjunto de datos.
- g. *Granularidad*: nivel de detalle con que se presentan los datos.
- h. Documentos anexos: información adjunta la cual tiene por objetivo complementar el contenido del recurso asociado. Los campos a llenar son:
 - *Título*: título definido para el documento asociado.
 - *Enlace de distribución*: URL del archivo.
 - *Descripción*: descripción del documento asociado.
 - *Tamaño*: volumen físico utilizado por el archivo, en unidades estándares. Se asigna de manera automática.
 - *Formato*: modos o medios principales de representación de los datos, normalmente identificados en la extensión del nombre del archivo. Se asigna de manera automática.

Aquellos conjuntos de datos que estén constituidos por uno o más archivos deben ser posibles de reconstituir a partir de los metadatos que se describen en la distribución del conjunto de datos.



En aquellos casos en que los datos son producto del procesamiento de otros conjuntos de datos, se debe indicar la procedencia de los datos que lo conforman.

6.2 Licencias

Se deben utilizar licencias que permitan el libre uso de la información publicada en el portal, es decir, que no tenga barreras para su reutilización y difusión. Algunos ejemplos son:

Licencias de contenido

- Attribution (CC BY)
- Attribution Share Alike (CC BY-SA)
- The GNU Free Documentation License

Licencias de datos

- Open Data Commons Public Domain Dedication and License (PDDL)
- Open Data Commons Attribution License
- Open Data Commons Open Database License (ODbL)
- Creative Commons CC0 Public Domain Dedication

6.3 Estructura del portal de datos abiertos

Todo portal de datos abiertos debe contener como mínimo los siguientes componentes:

- Un catálogo que permita visualizar el contenido general de los datos.
- Facilidades, herramientas y estructuras para explorar los diferentes elementos del catálogo, permitiendo el uso básico de los datos publicados.

Nota: El portal de datos abiertos deberían proveer acceso a los datos a través de interfaces o agentes automatizados para buscar y recuperar datos, permitiendo que estos sean utilizados por aplicaciones externas sin intermediación manual.

6.3.1 Catálogo

Los catálogos dirigidos al usuario de los conjuntos de datos publicados por la organización, dirigido para facilitar al usuario el acceso deben:

1. Contener un listado exhaustivo de los conjuntos de datos presentes en el repositorio de la organización (es decir, todo dato publicado en el repositorio debe poder ser accesible a través del catálogo).
2. Permitir la navegación, búsqueda y consulta simple de los datos contenidos en el repositorio, además si existen catálogos secundarios o específicos, estos deben estar enlazados con el catálogo central.

6.3.2 Exploración y uso básico de los datos

El portal de datos abiertos debe proveer al menos las funcionalidades de navegación y búsqueda de contenidos, permitiendo observarlos, analizarlos, imprimirlos y exportarlos, ya sea como datos o como referencia a los datos para ser utilizados desde otros sitios web.

Nota: Los portales de datos abiertos deberían incluir funcionalidades que permitan la realimentación de los usuarios tales como consulta y sugerencias.

7. EVALUACIÓN DE LA MADUREZ DEL PORTAL DE DATOS ABIERTOS

A continuación se presenta un modelo de madurez para portales de datos abiertos, el cual puede ser aplicado de manera complementaria a aquellos portales que aspiren a tener un nivel de madurez mayor que el nivel 1 o Básico, que se alcanza mediante la conformidad de todo lo establecido en los capítulos anteriores de esta norma.

Si la organización propietaria del portal de datos abiertos desea optar por incrementar su nivel de madurez, podrá hacerlo sometiéndose a una evaluación adicional donde se verifique la conformidad con los requisitos que se muestran a continuación.

7.1 Niveles del Modelo de Madurez para Portales de Datos Abiertos

Este modelo de madurez para portales de datos abiertos está conformado por 3 niveles. A continuación se explica cada uno de los niveles.

7.1.1 Nivel 1: Básico

El nivel 1 o Básico se asignará a aquel portal de datos abiertos que cumpla con todo lo establecido en los capítulos del 1 al 6 de esta norma.

7.1.2 Nivel 2: Notable

El nivel 2 o Notable se asignará a aquel portal de datos abiertos que cumpla con lo establecido en los capítulos del 1 al 6 de esta norma y que además cumpla con los requisitos establecidos para cada criterio en el nivel 2.

7.1.3 Nivel 3: Destacable

El nivel 3 o Destacable es el máximo nivel que puede recibir un portal de datos abiertos mediante la aplicación de este modelo de madurez. Lo obtendrá aquel portal que cumpla con todos los requisitos del modelo, en todos los criterios de todos los niveles.

7.2 Criterios del Modelo de Madurez para Portales de Datos Abiertos

Para este modelo se han establecido 5 criterios de calidad con base en los cuales será posible determinar el nivel de madurez de un portal de datos abiertos. Para cada criterio de calidad se ha establecido un requisito en cada uno de los niveles, tal y como se detalla a continuación:

7.2.1 Requisitos del criterio D/R – Disponibilidad y Reutilización

Para alcanzar los niveles 2 o 3 se debe cumplir con los siguientes requisitos:

Nivel	Requisito
Nivel 2: Notable	El portal debe permitir integrar distintos conjuntos de datos y otros elementos (como gráficos) para exportarlos de manera combinada con el fin de ser analizados desde otras herramientas.
Nivel 3: Destacable	Al menos uno de los conjuntos de datos publicados en el portal debe ser reutilizado como mínimo por una comunidad que desarrolle o haga uso de aplicaciones informáticas que consumen dichos datos abiertos.

7.2.2 Requisitos del criterio REL – Relevancia

Para alcanzar los niveles 2 o 3 se debe cumplir con los siguientes requisitos:

Nivel	Requisito
Nivel 2: Notable	El portal debe ser referenciado por parte de otros sitios de consulta, como para ser considerado una fuente relevante de consulta sobre el tema.
Nivel 3: Destacable	El portal debe ser utilizado como fuente primaria de consulta de datos abiertos dentro de su entorno temático y geográfico.

7.2.3 Requisitos del criterio REP – Reputación

Para alcanzar los niveles 2 o 3 se debe cumplir con los siguientes requisitos:

Nivel	Requisito
Nivel 2: Notable	Las fuentes de datos de cada conjunto de datos en el portal deben ser primarias o reconocidas en su ámbito.
Nivel 3: Destacable	Las fuentes de datos que alimentan el portal se deben actualizar con una frecuencia acorde con su temática y su entorno.

7.2.4 Requisitos del criterio G/I – Granularidad e Integración

Para alcanzar los niveles 2 o 3 se debe cumplir con los siguientes requisitos:

Nivel	Requisito
Nivel 2: Notable	Cuando se hace un análisis de un conjunto de datos, el portal debe sugerir referencias a datos del mismo portal que tienen relación y que podrían complementar el análisis.
Nivel 3: Destacable	Cuando se hace un análisis de un conjunto de datos, el portal debe sugerir referencias a datos externos al portal que tienen relación y que podrían complementar el análisis.

7.2.5 Requisitos del criterio V/U – Visualización y Usabilidad

Para alcanzar los niveles 2 o 3 se debe cumplir con los siguientes requisitos:

Nivel	Requisito
Nivel 2: Notable	El portal debe permitir visualizar la misma información en diferentes formatos.
Nivel 3: Destacable	El portal debe incluir aplicaciones o herramientas adicionales de análisis y consumo avanzado de los datos abiertos publicados.

8. CORRESPONDENCIA

Esta norma no posee normas equivalentes.

9. ABREVIATURAS Y ACRÓNIMOS

Nota: Las URLs referencian información complementaria y/o especificaciones.

- **API (Application Programming Interface):** Interfaz de comunicación entre componentes de software que ofrece un conjunto de llamadas a ciertas librerías de programación que dan acceso a ciertos servicios desde los procesos, consiguiendo la abstracción en la programación entre niveles inferiores y superiores del software.
- **CR05:** Datum horizontal para Costa Rica del 2005
- **CRTM05 (Proyección Transversal de Mercator para Costa Rica del 2005):** Sistema oficial de coordenadas para Costa Rica, según decreto N° 33797-MJ-MOPT. El cual está asociado al datum horizontal para Costa Rica CR05 y al elipsoide del Sistema Geodésico Mundial (WGS84).
- **CSV:** Comma Separated Values. <http://www.ietf.org/rfc/rfc4180.txt>
- **CSW:** Catalog Service for the Web/ Catálogo de servicios en Web

- **DCAT (Data Catalog Vocabulary):** Vocabulario RDF para la interoperabilidad de catálogos de datos. Su objetivo principal es la expresión de los catálogos de datos de gobiernos en un formato estándar usando RDF.
- **DCMI:** Dublin Core Metadata Initiative. <http://dublincore.org/documents/dces/>
- **GML:** Geography Markup Language. <http://opengis.net/gml/>
- **HTML:** Hyper Text Markup Language. <http://www.w3.org/html/>;
- http://www.w3.org/egov/wiki/Data_Catalog_Vocabulary
- **IGN:** Instituto Geográfico Nacional de Costa Rica
- **JPEG:** Join Photographic Expert Group. <http://www.jpeg.org/>
- **KML:** Keyhole Markup Language. <http://www.opengeospatial.org/standards/kml/>
- **MIME (Multipurpose Internet Mail Extensions):** Serie de convenciones o especificaciones dirigidas al intercambio a través de Internet de todo tipo de archivos (texto, audio, vídeo, etc.) de forma transparente para el usuario.
- **MPEG-7:** Moving Picture Expert Group. <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
- **OGC (Consortio Geoespacial Abierto):** Es una organización internacional que define los estándares abiertos e interoperables para la publicación de información geográfica en Web, propiciando acuerdos entre las diferentes empresas del sector. <http://www.opengeospatial.org/>
- **PNG:** Portable Network Graphics. <http://www.libpng.org/pub/png/>
- **RDF (Resource Description Framework):** Infraestructura para describir semánticamente recursos, es decir, dotar de sentido a lo que representamos para que las máquinas lo comprendan. RDF se puede representar en distintos formatos: XML, N3, Turtle, etc. <http://www.w3.org/RDF/>
- **RDFa (Resource Description Framework-in-attributes):** Forma de representar los datos estructurados visibles en las páginas Web mediante unas anotaciones semánticas, incluidas en el código e invisibles para el usuario, que permiten a las aplicaciones interpretar esta información y utilizarla de forma eficaz. <http://www.w3.org/TR/2008/WD-xhtml-rdfa-primer-20080620/>
- **REST (Representational State Transfer):** cualquier interfaz web simple que utiliza XML y HTTP, sin las abstracciones adicionales de los protocolos basados en patrones de intercambio de mensajes como el protocolo de servicios web SOAP.
- **SDMX:** Statistical Data and Metadata eXchange. www.sdmx.org
- **SIG:** Sistema de Información Geográfica.
- **SKOS (Simple Knowledge Organization System):** vocabulario RDF para la representación de sistemas de conocimiento semi-formales, tales como tesauros, taxonomías y esquemas de clasificación. SKOS se ha diseñado para facilitar la migración de los sistemas organizacionales existentes a la Web Semántica.
- **SPARQL (SPARQL Protocol and RDF Query Language):** tecnología de consulta de información desde bases de datos y otros orígenes de datos en sus estados primitivos a través de la Web. Se compone de un lenguaje de consulta estandarizado y de un protocolo con el que ofrecer un

servicio Web estándar (HTTP / SOAP), permitiendo realizar consultas a diversas fuentes de datos que almacenan los mismos nativamente en RDF o los presentan como tal.

- **SPSS**: Statistical Package for Social Science. <http://www.spss.com/>
- **SSL/TLS**: Secure Socket Layer/Transport Layer Security. <http://datatracker.ietf.org/wg/tls/charter/>
- **STATA**: Data Analysis and Statistical Software. www.stata.com
- **SVG**: Scalable Vector Graphics. <http://www.w3.org/Graphics/SVG/>
- **TAR**: Tape Archiv.
<http://www.freebsd.org/cgi/man.cgi?query=tar&sektion=5&manpath=FreeBSD+8-current>
- **URI**: Uniform Resource Identifier. <http://tools.ietf.org/html/rfc3986>
- **W3C (World Wide Web Consortium)**: es un consorcio internacional de reconocido prestigio donde las organizaciones miembro, el personal a tiempo completo y el público en general, trabajan conjuntamente para desarrollar estándares Web. www.w3c.org
- **WCS**: Web Coverages Services / Servicio Web de Coberturas
- **WFS**: Web Features Services / Servicio Web de Características
- **WGS84**: World Geodetic System 1984. http://earth-info.nga.mil/GandG/publications/tr8350.2/tr8350_2.html
- **WMS**: Web Map Services / Servicio Web de Mapas
- **XHTML**: eXtensible Hyper Text Markup Language. <http://www.w3.org/TR/xhtml1>
- **XML**: eXtensible Markup Language. <http://www.w3.org/TR/2000/REC-xml-20001006>

10. REFERENCIAS

- CTIC, Glosario Open Data, <http://datos.fundacionctic.org/glosario-open-data/>
- Instructivo Presidencial N° 005 del 12 Noviembre del 2012. <http://www.gobiernoabierto.cl/sites/default/files/gab.pres.ndeg005.pdf>
- Norma ISO / TC 211
- Calderón, C. y Lorenzo, S. Open Government: Gobierno Abierto. Algón Editores, 2010
- Open Data Navarra, http://www.navarra.es/home_es/Open-Data/
- Open Data Euskadi, <http://opendata.euskadi.net/w79-home/es>
- Oviedo, E., Mazón J.N., Zubcoff, J.J. Towards a data quality model for open data portals. In Computing Conference (CLEI), 2013 XXXIX Latin American. IEEE, 2013.
- Oviedo, E., Mazón J.N., Zubcoff, J.J. Quality and maturity model for open data portals. In Computing Conference (CLEI), 2015 XXXXI Latin American. IEEE, 2015.
- Oviedo, E., Mazón J.N., Zubcoff, J.J. Considering open data quality for measuring maturity of open data portals. CLEI Electronic Journal, 2015.
- W3C, Publishing Open Government Data, <http://www.w3.org/TR/gov-data>



Anexo B: Formulación cuantitativa del modelo de calidad presentado en (Oviedo et al 2013)

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante

En (Oviedo et al 2013) se propuso una formulación cuantitativa que permite, a partir de cálculos matemáticos simples, calcular el valor que se debe asignar a cada una de las seis medidas del modelo de calidad de datos abiertos que fue presentado en dicho artículo. Cabe recalcar que esas medidas de calidad posteriormente evolucionaron y se convirtieron en los cinco criterios de calidad que, de manera conjunta, sustentan el modelo de madurez para portales de datos abiertos que es producto de esta investigación.

A continuación se presenta la formulación cuantitativa supra indicada:

Medida de Calidad: DISPONIBILIDAD

Dependiendo del formato en que hayan sido publicados los datos en el portal, se asigna como calificación un valor absoluto (entre 0 y 1) considerando las siguientes opciones:

- Si el dato no está disponible, el valor de disponibilidad es 0.
- Si el dato está disponible, el valor es 0.33.
- Si el dato está disponible y en formato estructurado, el valor es 0.67.
- Si el dato está disponible en formato no propietario, el valor es 1.

Medida de Calidad: CAPACIDAD DE REUTILIZACIÓN

Su valor puede ser 0, 0.5 o 1. Es el resultado de la suma de los valores que otorga cada uno de los siguientes dos sub criterios:

- Se considera 0.5 si se propicia la interpretación de los datos mediante metadatos.
- Se considera 0.5 si se propicia la realización de análisis más amplios dentro del contexto de los datos, permitiendo vincular datos publicados con otros externos.

Medida de Calidad: RELEVANCIA

La relevancia se establece a partir de los valores obtenidos de las siguientes fórmulas:

D_i = Cantidad de veces que se ha accedido al conjunto de datos abierto i .

$$D_i = |acc(i)|$$

R_i = Cantidad de conjuntos de datos que referencian al conjunto de datos i .

$$R_i = |j| \forall j \xrightarrow{refers} i$$

W_i = Relevancia para el conjunto de datos i , se calcula como el número de accesos al conjunto de datos i más el número de referencias al conjunto de datos i . Existe una ponderación mediante la constante j con el fin de priorizar uno u otro término.

$$W_i = j \cdot D_i + (1 - j) \cdot R_i$$

W_p = Relevancia del portal p , teniendo un valor entre 0 y 1. Se calcula la media aritmética de los valores de W_i para cada uno de los conjuntos de datos que pertenecen al portal de datos abiertos.

$$W_p = \frac{\sum_{i=1}^N W_i}{N}$$

Medida de Calidad: REPUTACIÓN

La reputación de un conjunto de datos se refiere a la reputación del portal donde se encuentran los datos abiertos. Para su cálculo se utiliza el algoritmo *PageRank* (Brin y Page 1998) aplicado al portal. Este valor estará siempre en el rango entre 0 y 1.

Medida de Calidad: GRANULARIDAD

Un atributo de un conjunto de datos puede estar relacionado con diferentes niveles de detalle o granularidad. Por ejemplo, el atributo provincia puede estar relacionado con país y con ciudad. Entonces, siendo L la cantidad de niveles que puede tener cierto atributo a y L_a el nivel de detalle en el que se encuentra dicho atributo en el conjunto de datos, entonces la granularidad de a es:

$$G_a = \frac{|L| - L_a + 1}{|L|}$$

La cantidad de niveles (L) se obtiene del contexto. Por ejemplo, para el atributo Provincia, si el número de niveles totales puede ser 3 (país=3, provincia=2 y ciudad=1), su $G = (3-2+1)/3 = 0.67$.

La granularidad del conjunto de datos i es la media aritmética de la granularidad de sus atributos, siendo A la cantidad total de atributos del conjunto de datos i .

$$G_i = \frac{\sum_{a=1}^A G_a}{A}$$

Medida de Calidad: VISUALIZACIÓN

Mide la relación entre la cantidad de gráficas (n) y los conjuntos de datos (N):

$$V = \frac{|n|}{|N|}$$



Anexo C: Carta emitida por la Directora de Normalización de INTECO

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



Bibliografía

Universitat d'Alacant
Universidad de Alicante



Universitat d'Alacant
Universidad de Alicante



- Abadal, E., Ollé Castellà, C., Abad-García, F., & Melero, R. (2013). Políticas de acceso abierto a la ciencia en las universidades españolas. *Revista Española de Documentación Científica*, 36(2), e007.
- Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2014). Meloda, métrica para evaluar la reutilización de datos abiertos. *El profesional de la información*, 23(6), 582-588.
- Ananthakrishna, R., Chaudhuri, S. and Ganti, V. (2002). Eliminating Fuzzy Duplicates in Data Warehouses, Proc. 28th Int'l Conf. Very Large Databases (VLDB '02).
- Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*.
- Avison, D., Lau, F., Myers, M. y Nielsen, P.A. (1999). Action Research. *Communications of the ACM* 42 (1), 94-97.
- Batini, C., Cappiello, C., Francalanci, C., Maurino, A. (2009) Methodologies for data quality assessment and improvement. *ACM Comput. Surv.* 41(3)
- Bedini, I., Farazi, F., Leoni, D., Pane, J., Tankoyeu, I., Leucci, S. (2014). Open Government Data: Fostering Innovation. <http://www.jedem.org/article/view/329>
- Berners-Lee, T. (1996) WWW: Past, Present, and Future. *IEEE Computer* 29(10): 69-77 (1996).
- Bizer, C., Heath, T., Berners-Lee, T. (2009) Linked Data - The Story So Far. *Int. J. Semantic Web Inf. Syst.* 5(3): 1-22.
- Braunschweig, K., Eberius, J., Thiele, M., Lehner, W. (2012) The State of Open Data: Limits of Current Open Data Platforms http://www2012.org/proceedings/nocompanion/wwwwebosci2012_braunschweig.pdf
- Brin, S., Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks* 30(1-7): 107-117
- Cafarella, M. J., Halevy, A., Wang, Z. D., Wu, E., Zhang, Y. Webtables: Exploring the power of tables on the web. In Proc. of VLDB, 2008.
- Elmagarmid, A., Ipeirotis, P. and Verykios, V. (2007). Duplicate record detection: A survey. *TKDE*, 19(1):1-16.
- Franklin, M.J., Halevy, A.Y., Maier, D. From databases to dataspace: a new abstraction for information management. *SIGMOD Record* 34(4): 27-33 (2005)
- French, W.L., Bell, C.H. Jr. (1996). *Desarrollo organizacional* (quinta edición). Prentice-Hall



Genero, M., Cruz-Lemus, J., Piattini, M. Métodos de investigación en Ingeniería del Software. RA-MA Editorial, 2014.

Graves, A., & Hendler, J. (2013). Visualization tools for open government data. In Proceedings of the 14th Annual International Conference on Digital Government Research (pp. 136-145).

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. Synthesis lectures on the semantic web: theory and technology, 1(1), 1-136.

Heise, A., Naumann, F. (2012) Integrating open government data with stratosphere for more transparency. J. Web Sem. 14: 45-56

Hogge, B. (2010). Open Data Study. http://www.transparency-initiative.org/wp-content/uploads/2011/05/open_data_study_final.pdf

Höchtel, J. (2015). Institutionalising open data quality: processes, standards, tools. Workshop on Open Data Quality: from Theory to Practice (ODQ2015).

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. Information Systems Management, 29(4), 258-268.

Janssen, M., & Zuiderwijk, A. (2014). Infomediary business models for connecting open data providers and users. Social Science Computer Review, 32(5), 694-711.

Jetzek, T., Avital, M., Bjorn-Andersen, N. (2012) The Value of Open Government Data: A Strategic Analysis Framework. In Proceedings of SIG eGovernment pre-ICIS Workshop, Orlando, USA.

Mazón, J.N., Zubcoff, J.J., Garrigós, I., Espinosa, R., Rodríguez, R. Open business intelligence: on the importance of data quality awareness in user-friendly data mining. EDBT/ICDT Workshops 2012: 144-147

McTaggart, R. (1991). Principles of Participatory Action Research. Adult Education Quarterly, 41(3).

Mergarejo, E. P., Vergara, I. P., & Ruíz, Y. R. (2014). Modelos de madurez y su idoneidad para aplicar en pequeñas y medianas empresas. Ingeniería Industrial, 35(2), 146-158.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. The Journal of Strategic Information Systems, 24(1), 3-14.

Open Data Institute. (2015) Open Data Maturity Model: Assessing your open data publishing and use. <http://theodi.org/guides/maturity-model>



Organización Universitaria Interamericana (OUI). (2012) Modelo de Madurez de Open Data. <http://datos-abiertos.oui-iohe.org/el-proyecto>

Oviedo, E., Mazón J.N., Zubcoff, J.J. (2013) Towards a data quality model for open data portals. In Computing Conference (CLEI), 2013 XXXIX Latin American (pp. 1-8). IEEE.

Oviedo, E., Mazón J.N., Zubcoff, J.J. (2015) Quality and maturity model for open data portals. In Computing Conference (CLEI), 2015 XXXXI Latin American. IEEE.

Oviedo, E., Mazón J.N., Zubcoff, J.J. (2015) Considering open data quality for measuring maturity of open data portals. CLEI Electronic Journal (aceptado).

Padak, N., Padak, G. (1994). Guidelines for Planning Action Research Projects: <http://archon.educ.kent.edu/Oasis/Pubs/0200-08.html>

Paulk, M. C. (2009). A history of the capability maturity model for software. ASQ Software Quality Professional, 12(1), 5-19.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. Communications of the ACM, 45(4), 211-218.

Rahm, E., Do, H. (2000). Data cleaning: Problems and current approaches. IEEE Data Eng. Bull. 23(4):3-13.

Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. Communications of the ACM, 41(2), 79-82.

Rohloff, M., An approach to assess the implementation of business process management in enterprises. 2009.

Seaman, C.B. (1999). Qualitative Methods in Empirical Studies of Software Engineering. IEEE Transactions on Software Engineering, 25(4), 557-572.

Umbrich, J., Neumaier, S., & Polleres, A. (2015a). Towards assessing the quality evolution of Open Data portals. In ODQ2015: Open Data Quality: from Theory to Practice Workshop. Munich, Germany, <http://polleres.net/publications/umbretal-2015ODQ.pdf>

Umbrich, J., Neumaier, S., & Polleres, A. (2015b). Quality Assessment and Evolution of Open Data Portals. In Future Internet of Things and Cloud (FiCloud), 2015 3rd International Conference on (pp. 404-411).

van der Waal, S., Węcel, K., Ermilov, I., Janev, V., Milošević, U., & Wainwright, M. (2014). Lifting Open Data Portals to the Data Web. In Linked Open Data--Creating Knowledge Out of Interlinked Data (pp. 175-195). Springer International Publishing.



Vetr, A., Torchiano, M., Canova, L., Iemma, R., Morando, F. Open Data Quality Measurement Framework: Definition and Application to Open Government Data. 2014.

Vossen, G., & Hagemann, S. (2010). Unleashing Web 2.0: From concepts to creativity. Elsevier.

Wadsworth, Y. (1998). What is Participatory Action Research? Action Research International.

Wagner, A. Ben (2010). Open Access Citation Advantage: An Annotated Bibliography. University at Buffalo, New York.

Wagner, S., Lochmann, K., Winter, S., Goeb, A., Klaes M., Quality models in practice: A preliminary analysis. 3rd International Symposium on Empirical Software Engineering and Measurement (ESEM). 2009.

Wang, R.Y., Strong, DD.M. (1996) Beyond Accuracy: What Data Quality Means to Data Consumers. J. of Management Information Systems 12(4): 5-33

Xia J. (2012) Metrics to Measure Open Geospatial Data Quality. Issues in Science and Technology Librarianship, vol. 68.

