



**ESTUDIO DE LA REUTILIZACIÓN
DE DATOS PARA LA GENERACIÓN
DE NEGOCIOS**

Autores:

**ALBERTO ABELLA
MARTA ORTIZ-DE-URBINA CRIADO
CARMEN DE PABLOS HEREDERO**

ESTUDIO DE LA REUTILIZACIÓN DE DATOS PARA LA GENERACIÓN DE NEGOCIOS

Proyecto seleccionado dentro
de la convocatoria 2016
del Programa de Innovación
Abierta (PIA) de la Fundación
Cotec para la Innovación

*La contribución
del sector
público a la
economía del dato*

Este proyecto ha contado con el
apoyo técnico del departamento
de Economía de Cotec.

ÍNDICE

1. RESUMEN EJECUTIVO	8
2. INTRODUCCIÓN	11
2.1. España en el contexto europeo	11
2.2. El modelo de reutilización de datos	14
2.2.1. El proceso de publicación de datos	15
2.2.2. Los reutilizadores de datos	16
2.2.3. El impacto de los datos reutilizados	18
2.2.4. Modelos de negocio de datos abiertos en el contexto europeo	19
2.2.5. Consumo de datos en organizaciones reutilizando datos en el contexto europeo	22
2.3. Barreras a la reutilización a nivel europeo	22
2.4. Objetivo del informe	23
3. METODOLOGÍA	24
3.1. Metodología de estudio de los portales que publican datos	24
3.1.1. Modelo simplificado de madurez de portales que publican datos	24
3.2. Metodología de estudio de los juegos de datos publicados	25
3.3. Metodología de estudio de la reutilización de los datos publicados	26
4. DIAGNÓSTICO	28
4.1. Diagnóstico de los portales que publican datos	28
4.1.1. Actualización de los datos y disponibilidad de API	28
4.1.2. Sistema de gestión de datos	28
4.1.3. Portal de servicios desarrollados	29
4.1.4. Madurez de los portales según la metodología	29
4.2. Diagnóstico de los juegos de datos publicados	30
4.2.1. Distribución por grado de madurez de los portales que los publican	30
4.2.2. Distribución según la madurez de los portales	30
4.2.3. Distribución por temáticas de la NTI-RISP	31

4.2.4. Categorización por el estándar técnico utilizado	32
4.2.5. Categorización por los mecanismos de acceso necesarios para el acceso a los datos	33
4.2.6. Categorización por el contenido geográfico	34
4.2.7. Categorización por la frecuencia de actualización	34
4.2.8. Categorización por reusabilidad global	35
4.3. Diagnóstico de la reutilización de datos	36
4.3.1. Análisis del conocimiento sobre las entidades que reutilizan datos publicados	36
4.3.2. Análisis reutilización de datos por sectores de actividad	37
4.3.3. Análisis reutilización de datos por ámbito territorial	39
4.3.4. Análisis de los tipos de innovación por la reutilización de datos abiertos	39
4.3.5. Análisis de la disponibilidad de registro de los accesos	40
4.3.6. Análisis de las actividades de promoción del uso de datos abiertos	40
4.4. Diagnóstico de los servicios generados	41
4.4.1. Análisis de las temáticas de los servicios	41
4.4.2. Análisis de sostenibilidad y modelos de negocio de los servicios	41
4.4.3. Análisis de los autores de los servicios	42
4.4.4. Análisis de otras características de los servicios	42
4.4.5. Análisis de la creación de valor de los servicios	42
5. ESTIMACIÓN CUALITATIVA DE SERVICIOS INNOVADORES	44
5.1. Tipos de modelos de negocio identificados	44
5.1.1. Datos de los usuarios	44
5.1.2. Anuncios	44
5.1.3. <i>Freemium</i>	45
5.1.4. Recomendaciones contextuales	45
5.2. Otros mecanismos de sostenibilidad del uso de datos abiertos	45
5.3. Análisis de los servicios por el modelo de negocio	46
5.4. Comparativa de servicios generados vs datos disponibles	46
6. DAFO	47
7. CONCLUSIONES	48

7.1. Principales barreras	48
7.2. Otras barreras	48
7.2.1. Barreras estratégicas	49
7.2.2. Barreras técnicas	49
7.3. Medidas para superar las barreras	49
7.3.1. Asociación de publicadores de datos	50
7.4. Futuras líneas de trabajo	51
7.4.1. Estudio de la innovación empresarial a partir de la reutilización de datos abiertos	51
7.4.2. Métricas de reutilización de datos	52
7.4.3. Estudio de las condiciones de generación de nuevos negocios desde las fuentes de datos	52
7.4.4. Actualización continua de los datos	52
8. BIBLIOGRAFÍA	53
9. ANEXOS	55
9.1. Relación de portales de datos muestreados	55
9.2. Cuestionario a los reutilizadores	58
9.3. Respuesta sobre los tres usos más comunes que los reutilizadores dan a los datos de su portal	60
9.4. Respuestas a los tres juegos de datos que más se usan en su portal	62
9.5. Características básicas de un sistema de gestión de datos (DMS)	63
9.6. Clasificación de estándares técnicos según MELODA versión 4	65

ÍNDICE DE ILUSTRACIONES

Ilustración 1:	Mercado de datos abiertos por sector para la EU28+ 2020 (en millones de euros)	11
Ilustración 2:	Tamaño del mercado directo para la EU28+ 2020	12
Ilustración 3:	Clasificación de países en función de la madurez de los datos abiertos en la EU28+	13
Ilustración 4:	Madurez del uso de datos abiertos	14
Ilustración 5:	Ecosistema de reutilización de datos abiertos	15
Ilustración 6:	Categorías de reutilizadores	16
Ilustración 7:	Modelo de creación de valor de la reutilización de datos	18
Ilustración 8:	Clasificación mecanismos de impacto	18
Ilustración 9:	Modelo conceptual para la generación de valor a partir de los datos abiertos públicos (Open Government Data)	19
Ilustración 10:	Arquetipos de modelos de negocio en la cadena de valor de los datos abiertos	20
Ilustración 11:	Arquetipos de la cadena de valor de los datos abiertos por tipo entidad	21
Ilustración 12:	Consumo de datos por temáticas en empresas reutilizando datos abiertos	22
Ilustración 13:	Distribución de la madurez de portales de datos en España	29
Ilustración 14:	Distribución de juegos de datos por madurez de los portales que los publican	30
Ilustración 15:	Distribución de juegos de datos por temáticas NTI-RISP	31
Ilustración 16:	Distribución de juegos de datos por el estándar de almacenamiento utilizado	32
Ilustración 17:	Distribución de juegos de datos por el mecanismo de acceso utilizado	33
Ilustración 18:	Distribución de juegos de datos por el contenido geográfico de la información	34
Ilustración 19:	Juegos de datos por frecuencia de actualización	35
Ilustración 20:	Distribución de juegos de datos muestreados por categorías reusabilidad MELODA	35
Ilustración 21:	Distribución del conocimiento de los reutilizadores de datos	36
Ilustración 22:	Distribución de autores de los servicios	42

ÍNDICE DE TABLAS

Tabla 1:	Métrica para analizar el grado de madurez de los portales de datos	25
Tabla 2:	Tabla de temáticas del muestreo de datos publicados	32
Tabla 3:	Explicación del metadato asociado a un formato técnico	33
Tabla 4:	Conocimiento del tipo de reutilizadores de datos	37
Tabla 5:	Reutilizadores por sector de actividad	38
Tabla 6:	Ámbito de actuación de los reutilizadores de datos abiertos	39
Tabla 7:	Tipos de innovación por la reutilización de datos abiertos	40
Tabla 8:	Disponibilidad de los registros de acceso a los datos	40
Tabla 9:	Actividades de promoción del uso de datos abiertos	41
Tabla 10:	Creación de valor por la reutilización de los datos	43
Tabla 11:	Análisis DAFO	47

RESUMEN EJECUTIVO

La progresiva digitalización de nuestras ciudades (**ciudades inteligentes¹**) y de los objetos que nos rodean (**Internet de las Cosas², objetos llevables³**) dibuja una sociedad donde el tratamiento masivo de los datos abre la puerta a una nueva era.

Los **datos abiertos, datos masivos o innovación basada en datos**, se perfilan actualmente como las vías por las que la sociedad digitalizada resolverá desafíos basados en el conocimiento y la reutilización de datos. Desafíos anteriormente inabordables, como la sostenibilidad pública y medioambiental, el transporte, la salud o el empleo.

Las instituciones públicas y las empresas generan importantes cantidades de datos abiertos⁴ que pueden reutilizarse para crear valor en múltiples dimensiones: economía, sociedad, medio ambiente, etc.

El sector público ha sido desde el comienzo de la economía de los datos un importante contribuidor, tanto en forma de regulación y normativa, como en forma de incentivos que proporcionan los cimientos de esta nueva economía. Su papel más proactivo es la creciente publicación de datos abiertos. Con ellos, se busca el círculo virtuoso entre los datos publicados y, que, basados en ellos, se pueden lograr tres tipos de efectos:

1. Se incrementa la **confianza en la gestión pública** al estar disponibles los datos de la gestión.
2. Se **incrementa la eficiencia** mediante una utilización más fácil de los recursos públicos adecuados.
3. Se permite la aparición de nuevas aplicaciones y servicios que, con base en dichos datos y procesos que añaden valor, generan **un impacto social y económico**.

La reutilización de los datos abiertos permite crear valor para la sociedad como consecuencia de la oferta de nuevos productos y servicios digitales y la dinamización de la actividad económica y empresarial. Entre otras cosas, la reutilización de datos ha permitido la creación de nuevas empresas que con pocos recursos propios llevan a cabo modelos de negocio basados en la elaboración de productos y servicios enriquecidos con información de valor añadido (Marcos-Martín y Soriano-Maldonado, 2011; Abella, Ortiz-de-Urbina-Criado y De-Pablos-Heredero, 2014).

Sin embargo, en muchos casos, el sector público no tiene establecidas, o no es el objeto de su actividad, las rutinas organizativas que le permitan desarrollar la capacidad para extraer todo el valor de los datos abiertos. Y es aquí donde las empresas pueden intervenir, creando servicios de valor añadido que hagan la información más comprensible y asequible (Ferrer-Sapena, Peset y Aleixandre-Benavent, 2011; Lee y Kwak 2012; Abella et al., 2014: 583).

1. Más de 80 sólo en España.
2. Más de 20.000 millones de acuerdo a los informes de Cisco para 2020.
3. Más de 600 millones, actualmente, volcando información de forma continua.
4. Los datos abiertos son aquellos que son publicados con una licencia que permite su reutilización para cualquier propósito y por cualquier persona o entidad con el único requisito de realizar la atribución de la fuente.

Uno de los objetivos de este estudio es la formulación de directrices y recomendaciones que ayuden a la generación de negocios. Para ello, se han identificado los portales de datos abiertos en España y se ha analizado una muestra de los juegos de datos que tienen disponibles y de los servicios basados en ellos. Además, se ha enviado un cuestionario a los responsables de los portales para analizar algunas de las características de la reutilización de sus datos. En concreto, se ha hecho un diagnóstico sobre el conocimiento que tienen acerca de la reutilización de sus datos, sobre el tipo de innovación que se puede promover con ellos, sobre las actividades de promoción del uso de datos, sobre los servicios generados y sobre la creación de valor en torno a la reutilización de los datos. Todos estos diagnósticos han permitido realizar un análisis DAFO, a partir del cual se han incluido algunas reflexiones que pueden ayudar a construir futuras políticas públicas de gestión de datos.

El estudio realizado nos permite poner de manifiesto las siguientes evidencias y reflexiones sobre el ecosistema de reutilización de datos en España:

- **BASE SÓLIDA.** El sector público nacional se encuentra bien posicionado en el contexto europeo y mundial como publicador y promotor de la reutilización de datos (1er puesto en la UE), según el informe del portal paneuropeo de datos de 2016.
- **SECTOR AVANZADO DE PUBLICACIÓN DE DATOS.** El sector, en general, podría calificarse como avanzado (74% de los datos publicados en portales con características avanzadas), si bien, entre fuentes, existen diferencias muy importantes en las facilidades ofrecidas para la reutilización de sus datos.
- **BASE NORMATIVA.** En España existe una buena base normativa en cuanto a regulación de la reutilización de datos del sector público: adaptación de la directiva europea de 2013 (Ley 18/2015), la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información (Resolución 19 de febrero de 2013) y otras normativas nacionales, regionales y locales.
- **CONSUMO INTERNO.** Hoy por hoy, según los responsables de los portales de datos (el 44% reutilizan frecuentemente o siempre), los mayores consumidores de los datos publicados son las propias administraciones públicas, así como los mayores generadores de servicios (43%) y, en su opinión, las mayores mejoras por la reutilización de datos se producen para el sector público.
- **DESAJUSTE ENTRE DATOS PUBLICADOS Y SERVICIOS GENERADOS.** La mayoría de los datos publicados (52%) no contiene ninguna referencia geográfica y sólo el 13% incluye coordenadas y textos. Además, sus frecuencias de actualización son bajas (para el 95% su frecuencia de actualización es superior a una semana). Sin embargo, los servicios generados son mayoritariamente geolocalizados (72%) y en una parte significativa, en tiempo real (35%).

- **SOSTENIBILIDAD Y MODELOS DE NEGOCIO INCIPIENTES.** Definiendo sostenibilidad como la viabilidad económica de un servicio a medio plazo se encuentra que un 50% de los servicios generados tienen esta característica. Sólo la mitad de ellos (25%) poseen un modelo de negocio independiente mientras que el resto (25%) cuentan con una entidad que sufraga los costes por diversos motivos (marketing, presencia institucional, etc.).
- **GOBERNANZA INTERNA DE DATOS.** La actualización de datos es un atributo crítico para incrementar su valor. Las bajas frecuencias de actualización, y sus desajustes (entre lo anunciado y lo realmente publicado), son compatibles con la existencia de procesos manuales de actualización y publicación, lo que supone una barrera para la generalización del uso de los datos.
- **COORDINACIÓN.** No aparecen mecanismos que proporcionen una adecuada coordinación, tanto para la homogeneización de los datos publicados como para los estándares de acceso a las distintas fuentes. En este punto se identifica la necesidad de una entidad que aborde esta coordinación a nivel europeo.
- **FALTA DE CONOCIMIENTO SOBRE EL USO DE DATOS.** La recopilación de información sobre la reutilización es limitada y el conocimiento que se da a los datos publicados y su impacto en la sociedad es muy limitado.
- **HERRAMIENTAS INADECUADAS.** Un limitado uso de herramientas apropiadas (40% usan Sistemas de Gestión de Datos) está limitando el desarrollo de servicios (p. ej. falta de automatización del acceso).
- **FALTA DE ESTANDARIZACIÓN DE LOS MODELOS DE DATOS.** Los datos publicados no comparten estructuras. Es decir, el mismo conjunto de datos, como, por ejemplo, los presupuestos, publicados por distintas entidades presentan frecuencias de publicación, niveles de agregación y contenidos distintos entre distintas fuentes. Nuevamente, resultan necesarios ciertos mecanismos para facilitar la homogeneización, de forma que ésta dispersión no imponga barreras a la reutilización y, por tanto, a la generación de negocios.

INTRODUCCIÓN

2.1.

ESPAÑA EN EL CONTEXTO EUROPEO

Berends *et al.* (2017) han realizado un estudio para cuantificar el valor que se puede crear en la EU28+ a través de los datos abiertos. Entre los indicadores del valor creado por los datos abiertos que se encuentran se encuentra la cuota de mercado como porcentaje del PIB. Se estima que esta cifra es el 0,39% en 2017 y que alcanzará el 0,47% en 2020. Por otra parte, si se analiza el tamaño del mercado por sector, se observa que es la Administración Pública el sector que obtiene un mayor beneficio de la apertura de datos, siendo además los primeros y más importantes reutilizadores de estos datos (Berends *et al.*, 2017) (Ilustración 1).

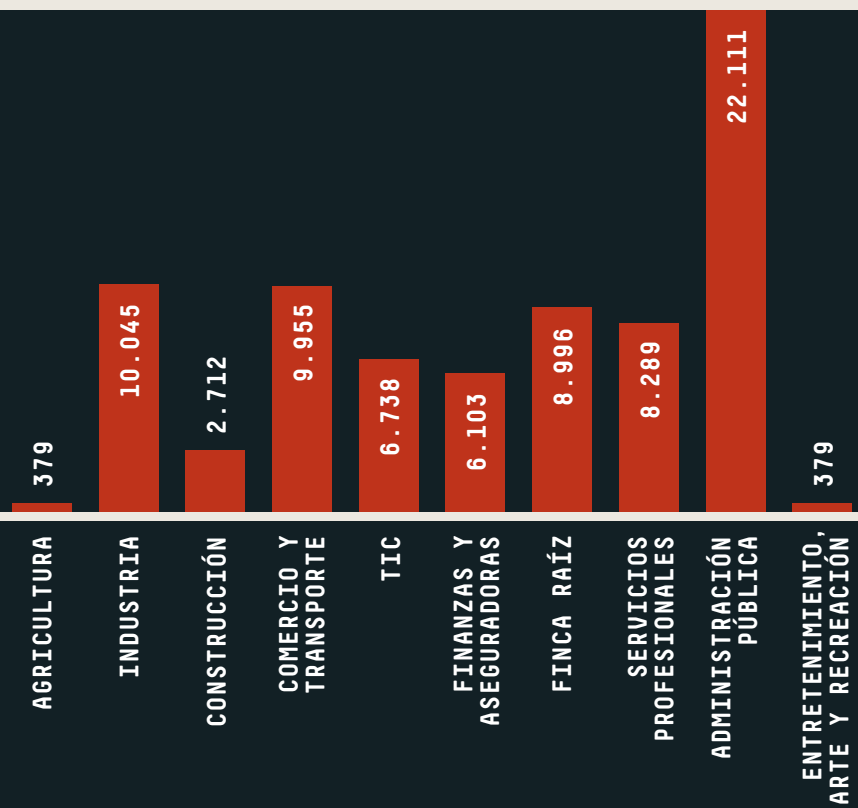


Ilustración 1

Mercado de datos abiertos por sector para la EU28+ 2020

(Cifras en millones de euros)

Fuente: Berends *et al.*, 2017

En cuanto a la creación de empleo, Berends *et al.* (2017) estiman que, a partir de la reutilización de datos abiertos, en 2017 se van a crear 80.500 empleos y se espera que, en 2020, haya aumentado en un 36,9%, es decir, 25.000 empleos más. Por otra parte, se observa también un efecto de los datos abiertos en el ahorro de costes en las Administraciones Públicas. Berends *et al.* (2017) estiman que, en el 2020, el ahorro que se puede lograr en EU28+ puede alcanzar los 1.700 millones de euros.

Por otra parte, según el estudio de Berends *et al.* (2017: 18-19) se espera que en el 2017 el tamaño del mercado directo de datos abiertos en la UE28+ sea de 59,7 miles de millones de euros, creciendo este tamaño un 36,9% entre 2016 y 2020, y llegando a alcanzar la cifra de 75,7 miles de millones de euros en 2020. Si consideramos el tamaño total acumulado de este mercado incluyendo los efectos directos, 325 miles de millones, y los indirectos, las estimaciones oscilan entre 1.138 y 1.229 miles de millones de euros. Sin embargo, el tamaño del mercado es diferente según el país. En países como Francia, Alemania, España y Reino Unido este mercado es muy importante (Ilustración 2).

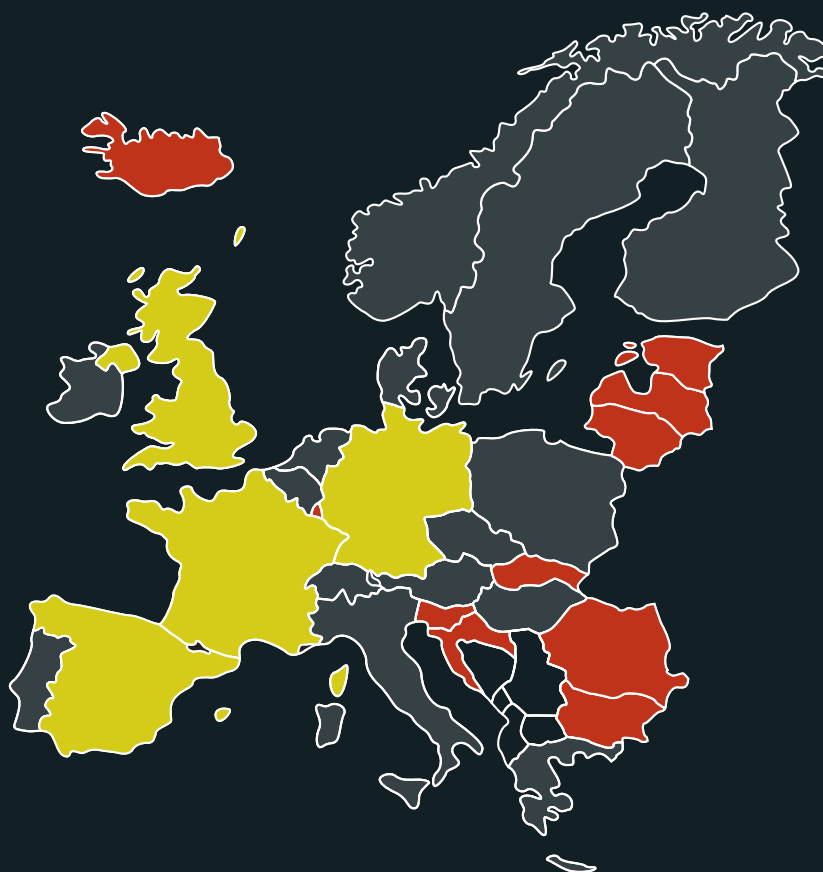
Ilustración 2

Tamaño del mercado directo para la EU28+ 2020

Clasificación



∞ Fuente: Berends *et al.*, 2017: 18-19



Según el informe del portal paneuropeo de datos (Carrara *et al.*, 2016), España es líder en cuanto a sus políticas de apertura de datos, por delante de Holanda y Francia, (2º y 3º respectivamente). También es líder en cuanto al impacto, en este caso por delante de Eslovaquia y Francia (2º y 3º) y se posiciona tercero en cuanto a la madurez de sus portales de datos, por detrás de Luxemburgo y Francia (1º y 2º). La clasificación global puede verse representada en la ilustración 4.

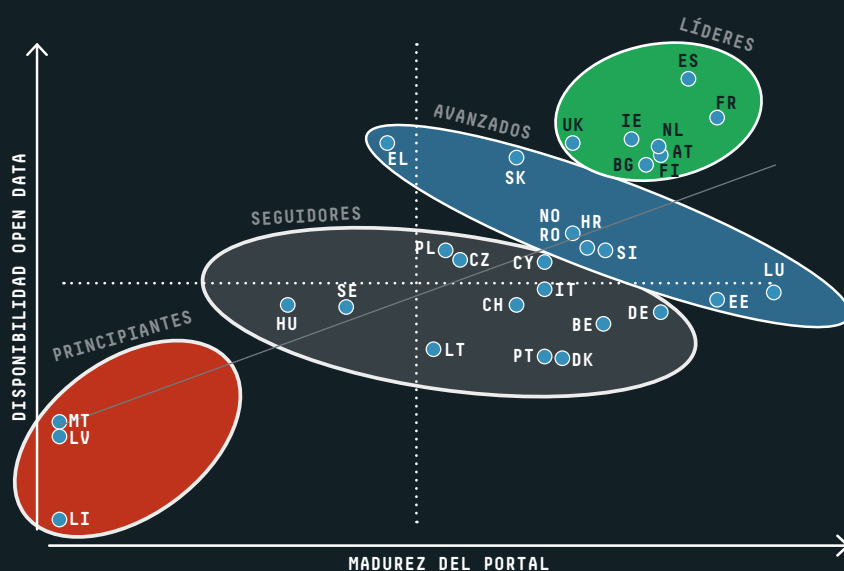
Carrara et al. (2016) analizan la madurez de los portales de datos abiertos para cada país de la EU28+. Los resultados obtenidos permitieron clasificar en cuatro grupos a los países estudiados (Ilustración 3). Se observó que tres países se situaban en la etapa inicial (*Principiantes*), 12 eran *Seguidores*, ocho estaban en la etapa de crecimiento rápido (*Avanzados*) y otros ocho en la etapa más avanzada, los *Líderes*. En este último grupo está España, que es la que tiene la cifra más alta de madurez (91,4%), seguida de Francia (85,8%), Irlanda (79,9%), Holanda (78,7%) y Austria (78,0%).



Ilustración 3:

Clasificación de países en función de la madurez de los Datos Abiertos en la EU28+

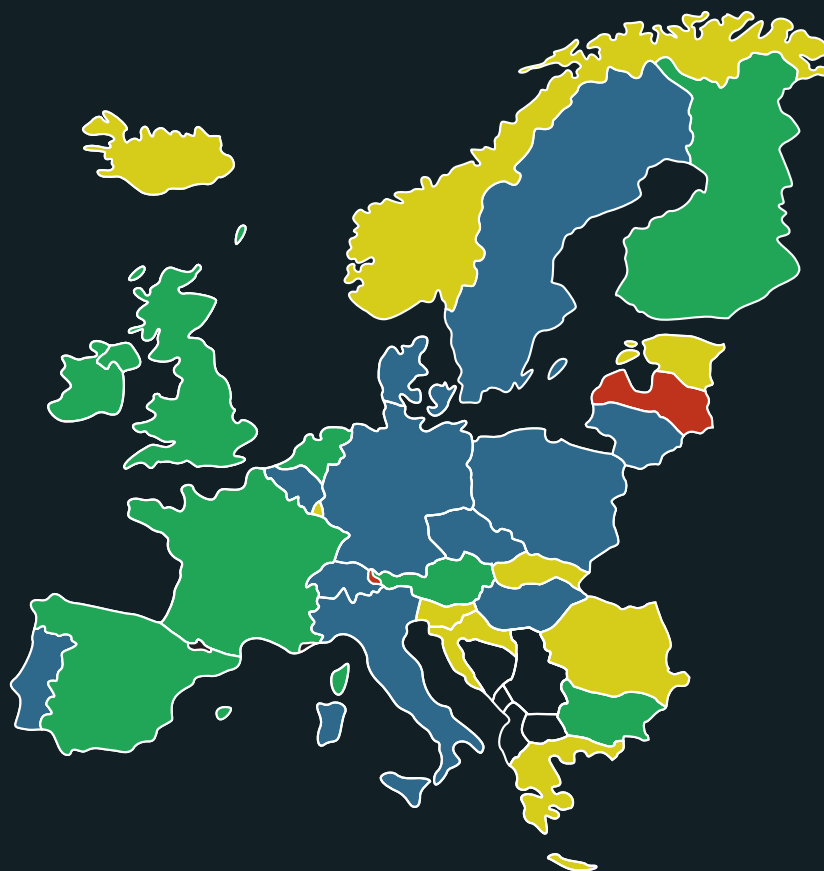
Fuente: Carrara et al., 2016



Además, según el informe del portal paneuropeo de datos (Carrara et al., 2016), si se desagrega la información para la valoración por países, se puede decir que España es líder en cuanto a sus políticas de apertura de datos (725 puntos sobre 790) por delante de Holanda y Francia, (2º y 3º), también es líder en cuanto al impacto (280 puntos sobre 300), en este caso por delante de Eslovaquia y Francia (2º y 3º) y es cuarta en cuanto a la madurez de sus portales de datos (220 sobre 250 puntos), por detrás de Luxemburgo, Francia y Estonia (1º, 2º y 3º). Si se considera la suma de todas las puntuaciones obtenidas, España se sitúa la primera en el ranking con un total de 1.225 puntos sobre 1.340⁵. La clasificación global puede verse representada en la ilustración 4.

5. Disponible en <http://www.europeandataportal.eu/en/dashboard#tab-detailed>

Ilustración 4

Madurez del uso de datos abiertos

∞ Fuente: Carrara et al., 2016: 61

2.2.**EL MODELO DE REUTILIZACIÓN DE DATOS**

La ilustración 5 muestra un modelo de reutilización de datos abiertos.

En este modelo se identifican, en primer lugar, los **portales de datos**, que son los recursos que ponen a disposición de la sociedad los datos abiertos.

En segundo lugar hay un ecosistema de agentes reutilizadores (explicados en el siguiente apartado) y en último lugar está la sociedad, entendiendo como tal todas aquellas personas, entidades y organizaciones que disfrutan de servicios basados en datos abiertos, bien sea de forma directa o indirecta.

En este sentido, la publicación de los datos se puede realizar de dos formas en función de su frecuencia de actualización. Los juegos de datos, que son grupos de datos estructurados recuperables en un enlace o instrucción única en su conjunto a una entidad única, con frecuencia de actualización mayor que una vez por minuto. Y de los chorros de datos (*dataJets*), que son grupos de datos estructurados recuperables en un enlace o instrucción única en su conjunto a una entidad única, con frecuencia de actualización menor que una vez por minuto.



Ilustración 5:
Ecosistema de reutilización de datos abiertos

Fuente: Elaboración propia

2.2.1. EL PROCESO DE PUBLICACIÓN DE DATOS

La publicación de datos abiertos se realiza siguiendo un proceso con varias etapas:

1. **Estrategia y diagnóstico.** En esta etapa se toma la decisión estratégica de contribuir con datos abiertos a la sociedad.
2. **Preparación de los datos.** En esta etapa han de prepararse una conexión lo más automatizada posible entre los sistemas internos que generan la información y los datos a publicar. También ha de tomarse la decisión de sobre qué datos publicar y, en consecuencia, se ponen en marcha los cambios organizativos que soporten la publicación continuada de datos.
3. **Publicación de datos.** En esta etapa se utiliza un recurso técnico (portal de datos) que los pone a disposición del público, incluyendo para su operación la organización y normativa que se ha desarrollado en la etapa anterior.
4. **Difusión.** El consumo de datos no se produce de forma espontánea, por lo que los publicadores de datos deben realizar actividades para que sus datos sean conocidos y utilizados por la sociedad, que genera valor añadido una vez incluidos en alguna solución o servicio.
5. **Evaluación.** En esta etapa se debe contrastar la demanda de los datos con las expectativas y analizar la retroalimentación de los reutilizadores, de forma que se revise las prioridades de publicación, los mecanismos técnicos y, en general, todo el esquema de publicación.

Cualquier deficiencia en estas etapas puede impactar de forma considerable, bien en el ritmo de publicación, bien en la reutilización efectiva de los datos.

2.2.2. LOS REUTILIZADORES DE DATOS

Los datos publicados en los portales pueden ser utilizados por diferentes agentes para crear productos o servicios para la sociedad. De esta forma, para poder analizar la reutilización de los datos abiertos hay que identificar a los agentes de este ecosistema, es decir, a los **reutilizadores**. Abella et al. (2017a) propone la siguiente clasificación de reutilizadores (Ilustración 6):

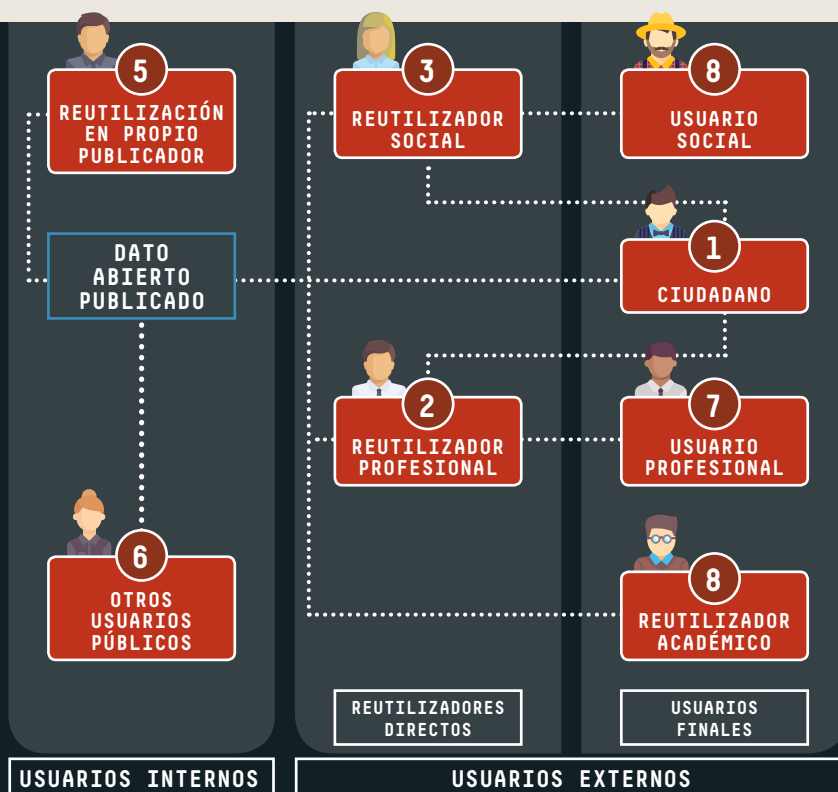


Ilustración 6:
Categorías de reutilizadores

Fuente: Abella et al., 2017a

La ilustración 6 representa el proceso de publicación y reutilización de datos por los distintos tipos de usuarios. Desde la fuente original (Open data publicado en la ilustración 6), los primeros reutilizadores son las entidades que publican la información (5 en la ilustración) y otras relacionadas (6 en la misma ilustración). Tras esto los reutilizadores directos (2 y 3 en la ilustración) toman los datos y generan servicios que consumen, preferentemente, los usuarios finales (1, 7 y 8 en la ilustración). Los reutilizadores académicos (4 en la ilustración) son un tipo especial que con frecuencia acuden a las fuentes de manera directa, como también hacen algunos de los ciudadanos (1 en la ilustración).

Los reutilizadores o usuarios internos, son aquellos que se encuentran, bien en el interior de las organizaciones que los publican o bien en organizaciones afines. Se corresponden fundamentalmente con el elemento *portal de datos* en la Ilustración 5: *Ecosistema de reutilización de datos abiertos*. Existen dos grupos:

- **La propia organización** que publica los datos (otras áreas o departamentos).
- **Otras entidades públicas** distintas de la que publica los datos

Los reutilizadores directos son aquellas organizaciones que generan los servicios y aplicaciones basados en los datos (y en otras fuentes o procesos que los enriquecen) y, en general, no son consumidores finales de los mismos. Se corresponden fundamentalmente con el elemento *ecosistema* en la Ilustración 5: *Ecosistema de reutilización de datos abiertos*. Hay dos tipos:

- **Reutilizadores profesionales** con ánimo de lucro (infomediarios, incluyendo también a desarrolladores individuales establecidos con ánimo de lucro).
- **Reutilizadores profesionales sin ánimo de lucro** (ONG, fundaciones, desarrolladores individuales sin ánimo de lucro, desarrolladores previos al lanzamiento de empresas y otras iniciativas sociales).

Finalmente, los reutilizadores indirectos son aquellos que consumen los datos en forma de servicios y aplicaciones, en algunos casos sin ser conscientes de ello. Se corresponden fundamentalmente con el elemento *sociedad* en la Ilustración 5: *Ecosistema de reutilización de datos abiertos*. Hay 4 tipos:

- **Ciudadanos individuales.**
- Investigadores y personal **académico** (incluidos estudiantes).
- **Usuarios sociales**, provenientes de organizaciones sin ánimo de lucro.
- **Usuarios profesionales**, provenientes de entidades con ánimo de lucro.

2.2.3. EL IMPACTO DE LOS DATOS REUTILIZADOS

El último paso es analizar el **impacto** que tiene la reutilización de los datos abiertos. Abella *et al.* (2017b) proponen un modelo de análisis del impacto de la reutilización de los datos abiertos (Ilustración 7). Los portales de datos son los que publican los juegos de datos a partir de los cuales se puede analizar el grado de reutilización de la información y su valor. La reutilización de estos datos permite crear productos y servicios innovadores, por ejemplo, aplicaciones para el ciudadano o para las empresas, cuya distribución y/o comercialización puede tener un impacto económico y social. Sin embargo, Abella *et al.* (2017b) señalan que todavía hay un bajo nivel de evidencias empíricas basada en datos reales del mercado que analice este impacto económico y social, con notables excepciones como los trabajos realizados por **Red.es** (Red.es 2011, 2012, 2015 y 2017).

Ilustración 7:
Modelo de creación de valor de la reutilización de datos

Fuente: Abella *et al.*, 2017b



En este sentido, uno de los problemas puede estar en la falta de información que tienen los portales de datos sobre la reutilización que se está haciendo de sus datos (Ver ilustración 21 en página 36).

El portal paneuropeo de datos propone una clasificación del impacto según se describe en la Ilustración 8.

Ilustración 8:
Clasificación mecanismos de impacto

Fuente: Carrara *et al.*, 2015



En este informe se dividen los efectos de los datos abiertos en tres dimensiones principales.

Rendimiento (performance). Evalúa los cambios en la eficiencia y calidad de los servicios gracias a la gestión del área de gestión de datos.

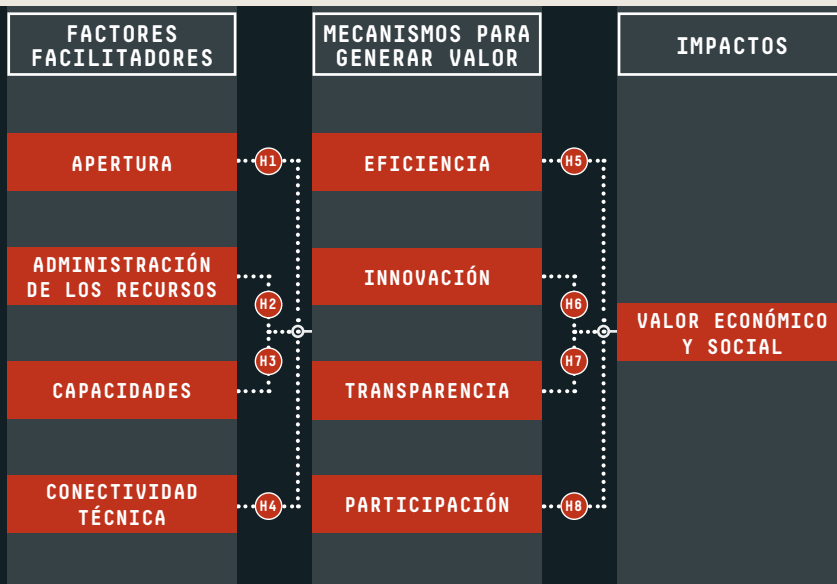
Economía. Evalúa el impacto de la apertura de datos y de las innovaciones que se pueden generar basados en ellos.

Social. Evalúa el impacto por la mejora de la transparencia y la rendición de cuentas.

Los dos últimos aspectos de impacto económico y social han sido así mismo evaluados de forma cuantitativa y cualitativa por (Jetzek et al., 2014a,b). Estos autores proponen un modelo de factores para evaluar el impacto económico y social de la reutilización de datos que se dibuja en la ilustración 9.

Ilustración 9:
Modelo conceptual para la generación de valor a partir de los datos abiertos públicos (Open Government Data)

Fuente: Jetzek et al., 2014a,b



En este modelo una serie de factores habilitadores influyen sobre los cuatro mecanismos identificados de generación de valor: aumento de la eficiencia, innovación, transparencia y participación ciudadana).

2.2.4. MODELOS DE NEGOCIO DE DATOS ABIERTOS EN EL CONTEXTO EUROPEO

El estudio de Berends et al. (2017) describe 5 arquetipos de modelo de negocio en la cadena de valor de los datos abiertos, basado en los identificados en el estudio de Dekers et al. (2006), según se visualiza en la ilustración 10.

Estos arquetipos se corresponden con distintas formas de creación de valor a lo largo de la cadena de valor de los datos abiertos. Los facilitadores (enablers) permiten y/o facilitan el acceso a los datos en cualquiera de las etapas de la cadena, y en general, no usan los datos ellos mismos.

Un ejemplo de este modelo es Libreforme.net (<http://libreforme.net>), una iniciativa para poner al alcance de los usuarios los datos del registro mercantil, que de otra manera, aunque son públicos hay que pagar para obtenerlos. Los proveedores (*suppliers*) generan datos para que sean consumidos en etapas posteriores. En esta categoría podrían encontrarse cualquiera de los portales de datos abiertos a nivel nacional y en concreto cualquiera de los que se encuentran en el punto 9.1. Los agregadores (*aggregators*) acumulan datos y pueden permitir el filtrado o selección de los mismos, así como la mezcla proveniente de múltiples fuentes y la extracción de valor desde su análisis. Un ejemplo de este modelo sería Datary, (<http://www.datary.io>). Ésta proporciona que proporciona servicios de comercialización de datos en empresas que poseyendo los datos, realizan una baja explotación de los mismos. Los enriquecedores (*enrichers*) utilizan los datos para añadir valor a productos o servicios existentes. Un ejemplo de este tipo sería BBVA Data & Analytics (<https://www.bbva.com>), la cual focaliza en la mejora de los procesos internos y la realización de análisis de datos. Finalmente, los desarrolladores utilizan los datos abiertos fundamentalmente para generar aplicaciones o servicios. En este último grupo podría ubicarse a Moovit (<http://moovitapp.com>), que genera un servicio de movilidad urbana en ciudades de todo el mundo. No obstante, hay que destacar que pocas organizaciones tienen un modelo de negocio ‘puro’ en el sentido que ocupen solo un eslabón de esta cadena de valor y que, lo más habitual, es que aborden varios simultáneamente.

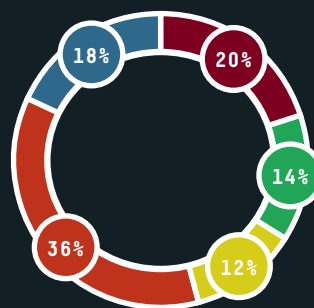
Ilustración 10:
Arquetipos de modelos de negocio en la cadena de valor de los datos abiertos

Fuente: Berends et al., 2017: 50

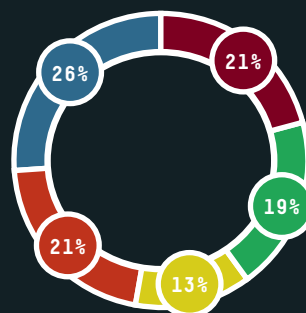


El muestreo del estudio de Berends *et al.* (2017) divide en tres sectores (privado, público y tercer sector, que incluye ONGs y entidades sin ánimo de lucro) a los utilizadores de datos y cualifica su papel en la cadena de valor de los datos abiertos. Los resultados se muestran en la ilustración 11.

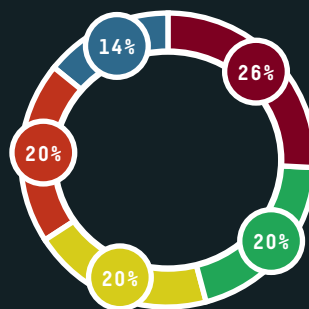
ARQUETIPOS [SECTOR PÚBLICO]



ARQUETIPOS [SECTOR PRIVADO]



ARQUETIPOS [TERCER SECTOR]



- AGREGADOR DE DATOS
- PROVEEDOR DE DATOS
- DESARROLLADOR
- ENRIQUECEDOR DE DATOS
- FACILITADOR DE DATOS

Ilustración 11:
Arquetipos de la cadena de valor de los datos abiertos por tipo entidad

∞ Fuente: Berends *et al.*, 2017

2.2.5. CONSUMO DE DATOS EN ORGANIZACIONES REUTILIZANDO DATOS EN EL CONTEXTO EUROPEO

Berends *et al.* (2017) presentan otros resultados (ilustración 12) de un análisis de 76 organizaciones, fundamentalmente del ámbito privado, que reutilizan datos.

En la gráfica se muestra el tipo de datos que consumen para su actividad desglosado por temáticas de acuerdo a la clasificación Eurovoc, acorde a la norma europea DCAT-AP. Los resultados se presentan en la ilustración 12.

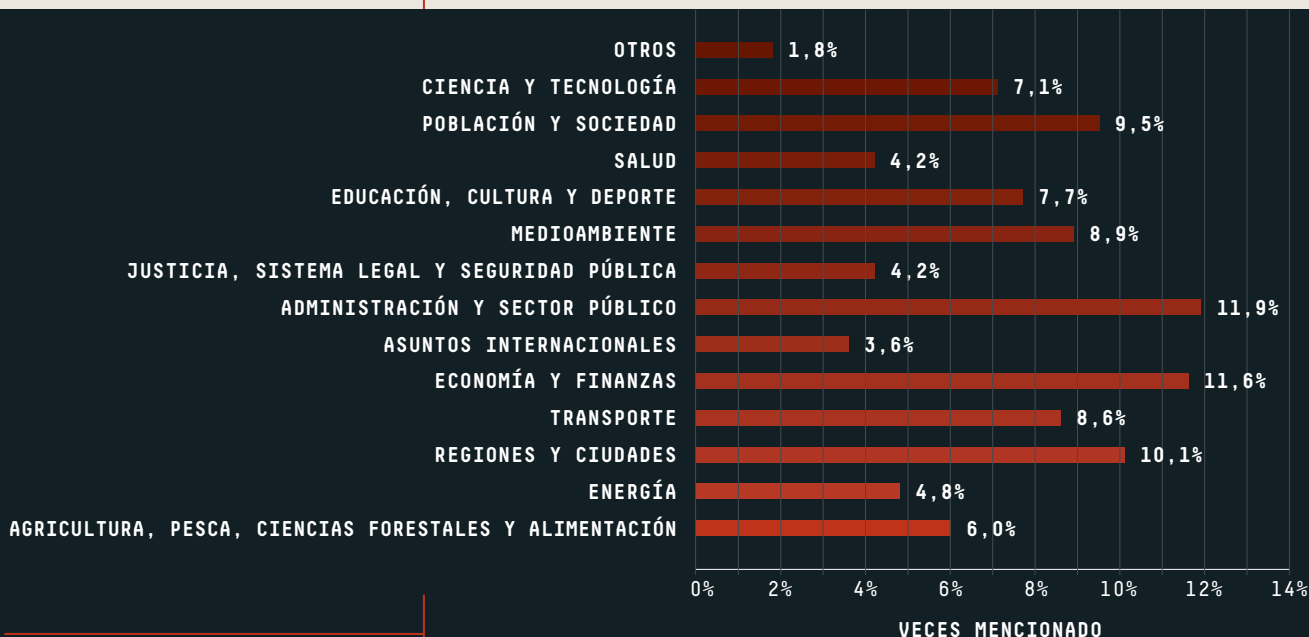


Ilustración 12:
Consumo de datos por temáticas en empresas reutilizando datos abiertos

Fuente: Berends *et al.*, 2017

2.3. BARRERAS A LA REUTILIZACIÓN A NIVEL EUROPEO

Estudios previos, como el proyecto europeo LAPSI II sobre la reutilización de datos del sector público (2015), han analizado las barreras legales y políticas a eliminar para fomentar la reutilización de información del sector público. Este estudio ofrece una serie de recomendaciones para aumentar la reutilización de los datos del sector público:

- Garantizar el mayor acceso posible a los recursos que proporcionan los datos.
- Limitar las restricciones a lo que se puede hacer con los datos.
- Proporcionar seguridad jurídica sobre los usos que se pueden hacer.

- Respetar la preferencia del usuario con respecto a, por ejemplo, el formato de los datos.
- Ayuda a reducir los costes de búsqueda.
- Dar a los reutilizadores voz en la toma de decisiones sobre cómo se liberan los datos.

También se ofrecen recomendaciones para el establecimiento de políticas como:

- Promocionar los derechos de acceso que incluyan la posibilidad de reutilización de la información en los estados miembros de la UE.
- Establecer la reutilización como un estándar en el marco europeo de interoperabilidad y las políticas de e-gobierno.
- Incluir una mayor comprensión y desarrollo del impacto del sector reutilizador en las legislaciones nacionales.
- Desarrollar y adoptar medidas preventivas para la protección de datos.

2.4.

OBJETIVO DEL INFORME

Teniendo en cuenta lo anterior, este informe realiza un estudio sobre la reutilización de datos en España con el objetivo de presentar el estado de la cuestión y de identificar directrices y recomendaciones que ayuden a fomentar el uso de los datos y a la generación de negocios. Para ello, se han identificado los portales de datos abiertos en España, se ha analizado una muestra de los juegos de datos que publican y se han analizado algunos de los servicios generados. Además, se ha realizado un cuestionario a los responsables de los portales para analizar algunas de las características y actividades en relación con sus datos abiertos. En concreto, se ha realizado un diagnóstico sobre conocimiento que tienen los portales sobre los reutilizados de datos, sobre el tipo de innovación que se puede realizar con estos datos, sobre las actividades de promoción del uso de datos, sobre los servicios generados y sobre la creación de valor de la reutilización de los datos. Todos estos análisis han permitido, por medio de un análisis DAFO, realizar un diagnóstico de las oportunidades y amenazas, y de las fortalezas y debilidades a partir del cual se han incluido algunas reflexiones que pueden ayudar a construir futuras políticas de gestión de datos para el fomento de la creación de negocios.

METODOLOGÍA

3.1. METODOLOGÍA DE ESTUDIO DE LOS PORTALES QUE PUBLICAN DATOS

De la totalidad de portales de datos (153) que están operativos en <http://mapa.datos.gob.es>, se han identificado 103 portales que publican datos abiertos (ver anexo 9.1). Este portal agrega juegos de datos a nivel nacional y está impulsado por la iniciativa Aporta perteneciente a los Ministerios de Hacienda y Función Pública, y de Energía, Turismo y Agenda Digital⁶.

Para cada portal se identificaron los siguientes valores:

- Disponibilidad de mecanismos de publicación de las actualizaciones de datos.
- Disponibilidad de un catálogo de recursos, número de juegos de datos disponibles y si el catálogo es descargable.
- Existencia de mecanismos de conexión directa con los datos (API).
- Disponibilidad de un portal donde se identifiquen servicios basados en los datos de los portales y número de servicios identificados.
- Utilización de una herramienta específica de publicación y reutilización de datos: sistema de gestión de datos o por sus términos ingleses DMS (ver anexo). 9.5.-Características básicas de un sistema de gestión de datos (DMS).
- Número de juegos de datos publicados.

3.1.1. MODELO SIMPLIFICADO DE MADUREZ DE PORTALES QUE PUBLICAN DATOS

Se ha definido un modelo de madurez, simplificado a partir del modelo definido por la iniciativa del portal paneuropeo de datos de Carrara *et al.* (2016), introduciendo los siguientes elementos a considerar:

- La población de juegos de datos que supere los 30 elementos.
- La disponibilidad de una fuente (canal sindicación rss o equivalente) con las actualizaciones de los datos.
- La disponibilidad de un interfaz de programación de aplicaciones (API) que permita el acceso automatizado a los datos por parte de usuarios externos.
- La utilización de un sistema de gestión de datos (DMS). A efectos de este trabajo, se han considerado las siguientes herramientas: DMS, CKAN, Socrata, DKAN y ESRI open data.
- La disponibilidad de un portal de aplicaciones con los servicios desarrollados basados en los datos publicados.

6. <http://aporta.es>

Con la información obtenida, se ha realizado un análisis de la madurez de los datos de los portales. Para ello, se ha desarrollado una métrica a partir de la propuesta de Carrara *et al.* (2016). La categorización de los resultados sigue el mismo grupo de categorías que la métrica de reutilización de datos MELODA (Abella *et al.*, 2014). Los aspectos analizados y su importancia se presentan en la tabla 1.

Tabla 1:
Métrica para analizar el grado de madurez de los portales de datos

Fuente: Elaboración propia a partir de Carrara *et al.*, 2016

CONCEPTO	CONCEPTO INFORME EUROPEO	PESO
Tener más de 30 juegos de datos	Difusión de los datos	20%
Tener una fuente con actualizaciones del catálogo	Usabilidad del portal	10%
Utilizar un sistema de gestión de datos (DMS)	Usabilidad del portal	15%
Disponibilidad de un interfaz de programación de aplicaciones (API) para interacción automatizada con los juegos de datos	Reutilización de los datos	25%
Portal de aplicaciones / servicios basados en datos abiertos	Reutilización de los datos	30%

3.2. METODOLOGÍA DE ESTUDIO DE LOS JUEGOS DE DATOS PUBLICADOS

Para el muestreo de los juegos de datos de los portales, se intentó la automatización de la descarga de datos desde los portales. Tras evaluar varias posibles alternativas (parte manual, parte automática) se determinó que no se disponía de recursos suficientes para que la automatización proporcionara suficientes resultados (más del 60% de los portales no tenían propiamente un sistema de gestión de datos DMS) por lo que un muestreo automatizado solo proporcionaría un número de datos interesante en el 40% de los casos. Por otra parte, la extensión a otros gestores de portales utilizados (CMS) era todavía más compleja por la mayor dispersión de herramientas utilizadas. Por tanto, el muestreo ha tenido que realizarse de forma manual.

En los 103 portales seleccionados se han identificado un total de 20.026 juegos de datos tras realizar un filtrado de los datos federados⁷ entre portales, que de no filtrarse serían contabilizados en más de una ocasión. Como ejemplo, la web datos.gob.es tiene más de 14.000 juegos de datos listados, si bien aquellos que no están federados de otras fuentes suman 1.200. De igual manera, el portal de la Comunidad de Murcia tiene cinco juegos de datos que no pertenecen a otras entidades y que no han sido contabilizados doblemente. Teniendo en cuenta la amplitud de la población objeto de estudio, se ha realizado un muestreo de 103 juegos de datos. De acuerdo con la herramienta *Surveysoftware* (<http://www.surveysoftware.net/sscalce.htm>) se ha seleccionado una muestra estadísticamente representativa de la población con un intervalo de 10 puntos al 95% de confianza.

7. Es un sistema múltiple de base de datos, en el cual, cada nodo en la federación mantiene su autonomía en los datos y define un conjunto de esquemas de exportación, a través de los cuales se hacen disponibles los datos a otros nodos.

Para cada juego de datos, se ha identificado la temática de acuerdo a la norma NTI-RISP (Resolución 19 de febrero 2013), así como cada una de las dimensiones de reusabilidad de datos descritas en la métrica MELODA (Abella et al., 2014)⁸:

- Licencia de reutilización.
- Estándar técnico en el que se presenta la información.
- Mecanismos de acceso.
- Modelo de datos utilizado.
- Contenido geográfico de la información.
- Frecuencia de actualización.
- Además del nombre del juego de datos y la URL de acceso.

3.3. METODOLOGÍA DE ESTUDIO DE LA REUTILIZACIÓN DE LOS DATOS PUBLICADOS

Se ha realizado una **encuesta a los portales de datos abiertos** durante el mes de marzo y principios de abril del 2017 (1-3-17 al 4-4-17), si bien los contactos para encontrar las personas responsables comenzaron en noviembre de 2016. De todos los portales analizados fue posible obtener contacto del responsable en 84 de ellos. Se obtuvieron 27 respuestas completas al cuestionario, lo que supone una tasa de respuesta del 32%.

Cada uno de ellos tuvo que responder un cuestionario de 11 preguntas sobre la reutilización de datos en su portal de datos, su conocimiento de los reutilizadores y sus políticas de promoción del uso (Anexo 9.2).

Por otra parte, se analizó la posibilidad de pasar un cuestionario a los reutilizadores de datos, pero los responsables de los portales de datos no tenían información suficiente sobre las organizaciones que reutilizan sus datos por lo que no fue posible realizar este análisis. Como alternativa se ha realizado un **análisis directo de las aplicaciones y servicios** que los propios portales de datos proporcionan como acreditados reutilizadores de datos, identificando desde este punto a los autores y extrayendo datos de los portales corporativos de cada uno de ellos.

Se han muestreado 59 aplicaciones y servicios de los 41 portales que tenían servicios inventariados basados en datos abiertos, elegidos de forma aleatoria. Lo que supone una muestra estadísticamente significativa para un intervalo de 12 puntos con un nivel de confianza del 95%, siguiendo la misma aproximación y herramienta que en el punto anterior.

8. La descripción completa de la métrica está disponible en el enlace <http://www.meloda.org/full-description-of-meloda/> con licencia Creative Commons 3.0 attribution

Los datos extraídos para cada uno de estos servicios han sido:

- La temática del servicio de acuerdo a la clasificación DCAT-AP (European Union, 2017).
- Las características de geolocalización.
- La característica de tiempo real del servicio.
- El tipo de autor de la aplicación utilizando la clasificación de reutilizadores propuesta en la ilustración 6 y descrita en el punto 2.2.2.
- El mecanismo de sostenibilidad en el caso de tenerlo. La sostenibilidad describe la viabilidad económica del servicio a medio plazo, bien porque haya una entidad que soporte los costes o bien porque el servicio tenga mecanismos propios de generación de ingresos. Por ejemplo, en el caso de un servicio corporativo, la entidad que lo publica asume sus costes pese a no tener ingresos directamente relacionados.
- Si tenía algún modelo de negocio que incluya una fuente de ingresos diferente a la de la entidad que publica los datos y en su caso de qué tipo.

DIAGNÓSTICO

4.1. DIAGNÓSTICO DE LOS PORTALES QUE PUBLICAN DATOS

El primer elemento analizado son las fuentes de datos (portales) que, desde el sector público, pueden alimentar y fomentar la economía basada en datos.

4.1.1. ACTUALIZACIÓN DE LOS DATOS Y DISPONIBILIDAD DE API

Del muestreo de 103 portales se detectan los siguientes datos:

Sólo el **28 % de los portales** tienen un canal que permita saber cuándo un juego de datos ha sido **actualizado** y sólo el **46%** disponen de un mecanismo automatizado de acceso a los datos (**API** o punto de interrogación semántica).

Es decir, el 72% de los portales han de ser muestreados de forma periódica por sus reutilizadores, con los problemas que ello comporta de incremento de costes para los mismos y de posible falta de utilización de los datos publicados. Además, teniendo en cuenta que la tendencia es que los datos se actualicen de forma cada vez más frecuente, esto conllevaría multitud de trabajo para los reutilizadores que requieran un muestreo periódico suficiente y para el propio portal que tendrá que estar dando respuesta a peticiones de información⁹.

El 54% de los portales no tiene un mecanismo automatizado de acceso (API), lo que supone un problema similar, ya que obliga a los reutilizadores a descargar todos los datos de un juego de datos, aun cuando solo necesitaran parte de ellos y a procesarlos para poder extraer los datos que realmente necesitan.

Finalmente, desde el punto de vista del responsable de publicación de datos, la descarga directa constituye un obstáculo ya que suele estar asociado a accesos anónimos a los recursos lo que impediría un análisis completo de los reutilizadores y los usos que se hacen de los datos.

Menos de la mitad de los portales de datos posee una sección donde se muestren los servicios creados con sus datos

4.1.2. SISTEMA DE GESTIÓN DE DATOS

En cuanto al sistema utilizado, el **40%** utilizan propiamente un sistema de gestión de datos (*CKAN, Socrata, DKAN, ESRI open data*), mientras que el **43%** utiliza una solución no específica (como algunos gestores de contenidos como *Joomla, Wordpress, Liferay*, etc.). Por otra parte, en el **17%** de los casos no ha sido posible identificar el gestor de la plataforma, no perteneciendo a los anteriores grupos, siendo normalmente desarrollos a medida.

El hecho de que no se estén utilizando soluciones específicas de gestión de datos (DMS) (Ver anexo 9.5) compromete la escalabilidad de estos sistemas. No así en cuanto a accesos directos, pero sí en cuanto a sus capacidades de gestión y de extracción de datos desde las organizaciones que los publican. De entre los portales en que no se utiliza una solución específica para

9. En este recurso <http://meloda.org/feed> están consolidados todos los canales con las novedades (nuevos juegos de datos) de los portales analizados en este informe, de tal forma que cualquier novedad en cualquier de ellos (lógicamente solo para aquellos que tienen canal de novedades) pueden ser accesibles en una única fuente.

publicar datos, sólo en un **23%** de los casos existe un acceso automatizado a los datos (API o punto de consulta semántico). Además, en estos casos, la documentación para facilitar el uso del API no suele estar al mismo nivel, tanto en profundidad como en actualización, comparado con aquellos sistemas que poseen una solución específica de gestión de datos y, por tanto, sus capacidades de interacción y de desarrollo de servicios, están más limitadas.

4.1.3. PORTAL DE SERVICIOS DESARROLLADOS

En cuanto al desarrollo de servicios tampoco llegan a la mitad (**40%**) los portales que tienen información sobre los servicios desarrollados basados en sus datos. Es decir, aquellos que tienen constancia de que se han creado, y hacen difusión de los servicios desarrollados a partir de los datos publicados.

4.1.4. MADUREZ DE LOS PORTALES SEGÚN LA METODOLOGÍA

La Ilustración 13 muestra la gráfica con el porcentaje de portales en función de su madurez para la publicación de datos. Para ello, se ha utilizado la metodología descrita en el punto 3.1. La ilustración muestra que, para cada intervalo de grado de madurez definido, el número de portales es muy similar, pese a que las características apreciadas en el modelo son muy diversas.

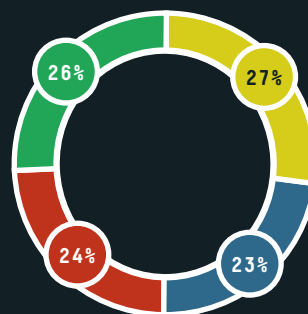
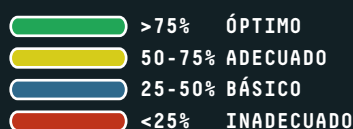


Ilustración 13:
Distribución de la madurez de portales de datos en España

Fuente: Elaboración propia

Del análisis de los datos descargables se han detectado importantes divergencias en el número de juegos de datos disponibles, desde fuentes que tienen menos de una decena de juegos de datos (26), hasta otras que tienen más de 1.000 (6). También se ha detectado que existen importantes diferencias en la estructuración de los datos. Por ejemplo, es frecuente que algunos portales separen como juegos de datos distintos el mismo conjunto de datos en los diversos años de los que dispone información. Incluso en esos años se puede volver a dividir por categorías aumentando de esta forma el número final de juegos de datos contabilizados.

4.2. DIAGNÓSTICO DE LOS JUEGOS DE DATOS PUBLICADOS

4.2.1. DISTRIBUCIÓN POR GRADO DE MADUREZ DE LOS PORTALES QUE LOS PUBLICAN

Si se suman los juegos de datos localizados en los portales que se encuentran en cada intervalo de madurez, se observa (ilustración 14) que el 91% de los juegos de datos se sitúan en portales con grados de madurez óptima o adecuada. Es decir, que los portales más avanzados son aquellos que además agregan mayor número de juegos de datos.

4.2.2. DISTRIBUCIÓN SEGÚN LA MADUREZ DE LOS PORTALES

La ilustración 14 refleja la suma de juegos de datos que se encuentra en los portales en función de su grado de madurez definido en el punto 3.1.1. Puede observarse que, buena parte de éstos se localizan en la categoría más avanzada (74%), ocupando esta sólo un 25% de los portales como se muestra en la ilustración 13. A continuación le siguen los portales con una madurez adecuada (27% del total de portales muestreados) y que representan el 17% de los juegos de datos disponibles. Esto implica que los portales más poblados de datos son además los más avanzados y los que más facilidades incluyen para la reutilización.

La tendencia en los portales de datos más avanzados es la consolidación de juegos de datos:

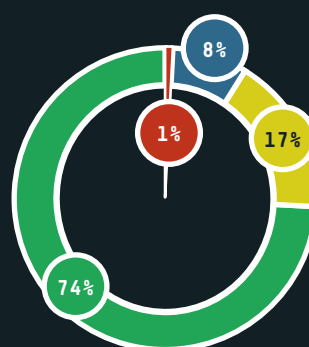
- Para simplificar las operaciones para los reutilizadores profesionales.
- Para reducir la gestión interna.
- Para reducir el coste en el caso de algunas herramientas de mercado.



Ilustración 14:

Distribución de juegos de datos por madurez de los portales que los publican

∞ Fuente: Elaboración propia



4.2.3. DISTRIBUCIÓN POR TEMÁTICAS DE LA NTI-RISP

La ilustración 15 muestra la distribución de juegos de datos en función de las temáticas definidas en la norma NTI-RISP (Resolución de 19 de febrero de 2013), destacando **empleo (14,6%)** y **hacienda (13,6%)**, donde se incluye todo el reporte de la ejecución de los presupuestos, seguido del **9,7%** de la temática de **urbanismo e infraestructuras**.

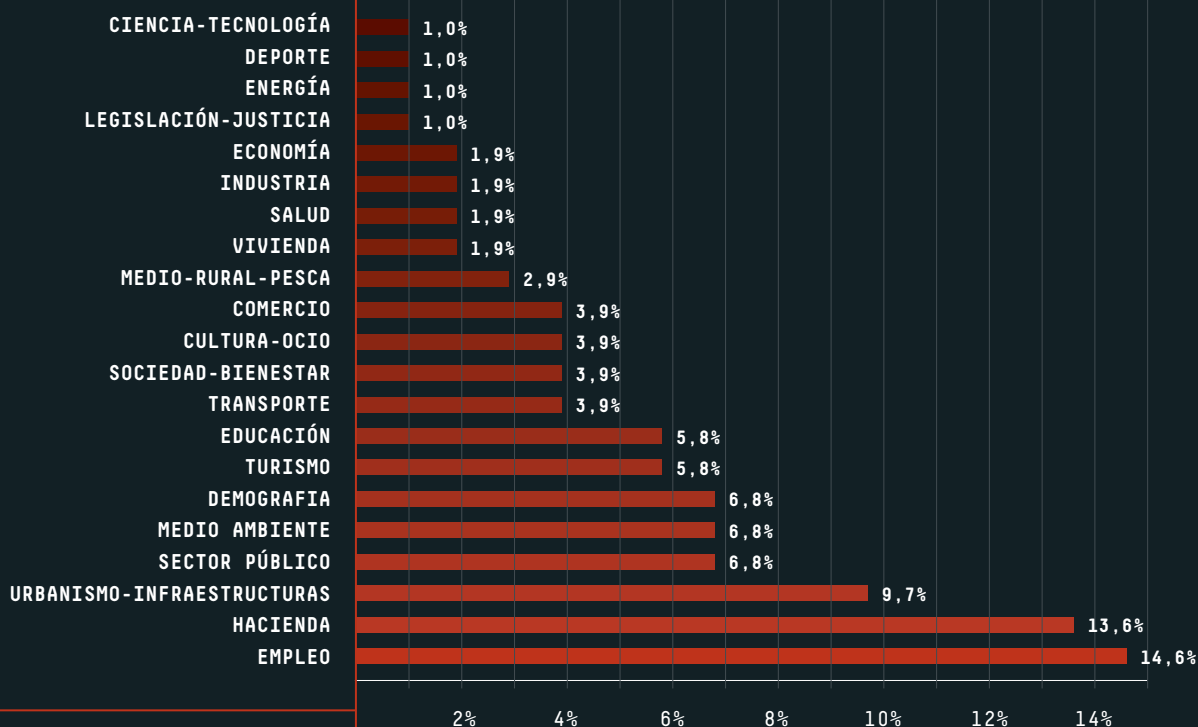


Ilustración 15:

Distribución de juegos de datos por temáticas NTI-RISP

Fuente: Elaboración propia

En la tabla 2 se incluyen los porcentajes obtenidos del muestreo, donde están incluidas todas las categorías definidas, salvo seguridad. Estas cifras se pueden deber, posiblemente, al tamaño muestral y al bajo porcentaje de informaciones publicadas en España sobre esta temática, que contrasta con otros países (como EE. UU.), donde se publica de forma proactiva y en tiempo real. Por ejemplo, el portal de datos de la ciudad de Chicago <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>, donde se obtiene información en tiempo real sobre delitos cometidos en la ciudad.

TEMÁTICA	%	TEMÁTICA	%
Empleo	14,6%	Transporte	3,9%
Hacienda	13,6%	Medio-rural-pesca	2,9%
Urbanismo-infraestructuras	9,7%	Economía	1,9%
Demografía	6,8%	Industria	1,9%
Medio Ambiente	6,8%	Salud	1,9%
Sector público	6,8%	Vivienda	1,9%
Educación	5,8%	Ciencia-Tecnología	1,0%
Turismo	5,8%	Deporte	1,0%
Comercio	3,9%	Energía	1,0%
Cultura-ocio	3,9%	Legislación-justicia	1,0%
Sociedad-bienestar	3,9%		

Tabla 2:

Tabla de temáticas del muestreo de datos publicados

Fuente: Elaboración propia

4.2.4. CATEGORIZACIÓN POR EL ESTÁNDAR TÉCNICO UTILIZADO

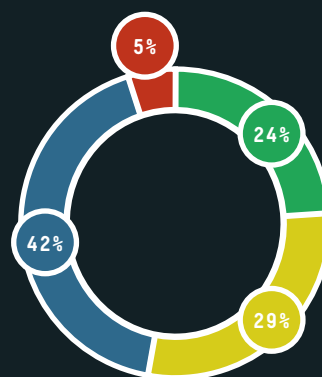
La ilustración 16 muestra la distribución de los juegos de datos en función de los estándares técnicos (Ver anexo 9.6) en que se publican los datos, donde sólo es destacable el 5% de juegos de datos que todavía se presentan en portales de datos abiertos y que en realidad no son reutilizables por su formato técnico (en la mayoría de los casos son pdf en formato imagen o equivalentes). Esto supone una barrera a la reutilización, no tanto desde el punto de vista técnico, sino desde el punto de vista económico, ya que, en general, la recuperación de la información desde este tipo de formatos ocasiona un gasto de recursos superior al valor que proporciona.

- █ NIVEL 1
Estándar cerrado no reutilizable
- █ NIVEL 2
Estándar cerrado reutilizable y estándar abierto no reutilizable
- █ NIVEL 3
Estándar abierto
- █ NIVEL 4
Estándar abierto con metadata

Ilustración 16:

Distribución de juegos de datos por el estándar de almacenamiento utilizado

Fuente: Elaboración propia



Los metadatos a los que se hace referencia en el nivel 4 son datos descriptivos de los datos aportados. En la siguiente tabla se incluye un ejemplo.

En la primera columna se sitúan los valores, pero se necesita un contexto para entender su significado, mientras que en la segunda los metadatos <temperatura localidad> ayudan a dar ese contexto para la utilización del dato.

Tabla 3:

Explicación del metadato asociado a un formato técnico

Fuente: Elaboración propia

Estándar técnico sin metadatos	Estándar técnico con metadatos
P. ej. formato csv	P. ej. formato XML
4°C, 6°C	<temperatura Madrid> 4°C </temperatura Madrid> <temperatura León> 6°C </temperatura León>

4.2.5. CATEGORIZACIÓN POR LOS MECANISMOS DE ACCESO NECESARIOS PARA EL ACCESO A LOS DATOS

Los mecanismos de acceso condicionan las herramientas que podrán utilizar los reutilizadores para el acceso automatizado a los datos. Sin embargo, es preferible disponer de API en todos los casos. Los lenguajes de interrogación, como SPARQL, permiten la consulta externa de los datos publicados. En este caso, el lenguaje SPARQL permite que las búsquedas recuperen datos de varias fuentes si éstas están enlazadas y comparten características técnicas comunes.

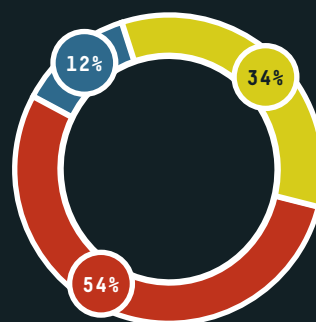
El 12% de juegos de datos requieren de una interacción con la web (no automatizada) para la extracción de los datos (Ilustración 17). De nuevo este porcentaje supone una barrera económica al acceso a los datos.

- NIVEL 2
Acceso por URL con registro o con interacción web
- NIVEL 3
Acceso a través de una URL única a todos los datos del juego de datos
- NIVEL 5
Acceso vía API o lenguaje interrogación (p. ej. SPARQL)

Ilustración 17:

Distribución de juegos de datos por el mecanismo de acceso utilizado

Fuente: Elaboración propia



4.2.6. CATEGORIZACIÓN POR EL CONTENIDO GEOGRÁFICO

La disponibilidad de contenido geolocalizado puede incrementar el interés de los datos publicados según se analizó en trabajos anteriores (Abella et al., 2015)¹⁰. Si bien es cierto que no todas las informaciones publicadas pueden ser geolocalizables, sí resulta destacable que más del 50% no presente ningún tipo de información geográfica y que solo el 16% tenga unas coordenadas geográficas asociadas al contenido publicado (ilustración 18).

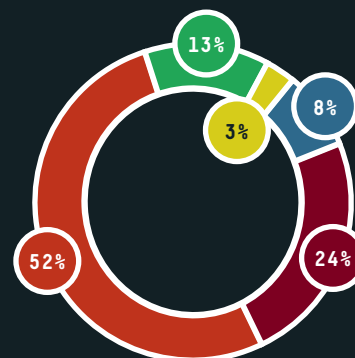
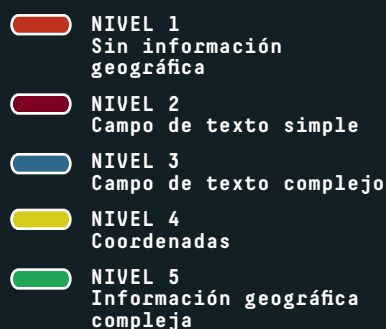


Ilustración 18:

Distribución de juegos de datos por el contenido geográfico de la información

∞ Fuente: Elaboración propia

4.2.7. CATEGORIZACIÓN POR LA FRECUENCIA DE ACTUALIZACIÓN

La ilustración 19 muestra la frecuencia de actualización de los juegos de datos muestreados. Resulta especialmente destacable que ninguno e los 103 esté por debajo de la actualización en horas o sea directamente chorro de datos (*dataJets*). Es decir, las categorías 4 y 5 de la ilustración 19 suman un 0% en el muestreo realizado. Y también es destacable el bajo porcentaje de las categorías 2 y 3, que escasamente alcanzan el 5%.

Al igual que sucede con la geolocalización de la información, no todas las informaciones son susceptibles de ser actualizadas en periodos tan frecuentes, o realizarlo ocasionaría un coste no asumible. Sin embargo, sí es cierto que los porcentajes detectados muestran un panorama donde la información es mayormente estática. Además, tampoco es infrecuente detectar juegos de datos donde no se cumplen las frecuencias de actualización anunciadas (P. ej. especificar que la frecuencia de actualización es mensual y no haberse actualizado en los últimos 8 meses).

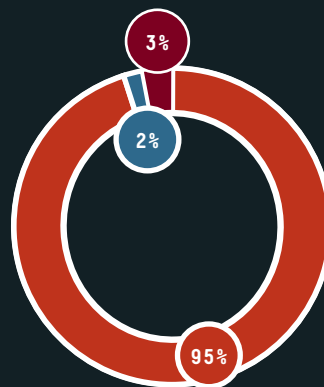
10. En este estudio se encontró que todas las informaciones (100%) más demandadas en una serie de ciudades inteligentes de Europa (19) siempre estaban geolocalizadas.

- NIVEL 1
Más de una semana
- NIVEL 2
1-6 días
- NIVEL 3
1-23 horas
- NIVEL 4
1-60 minutos
- NIVEL 5
< minuto (data)

Ilustración 19:

Juegos de datos por frecuencia de actualización

∞ Fuente: Elaboración propia



4.2.8. CATEGORIZACIÓN POR REUSABILIDAD GLOBAL

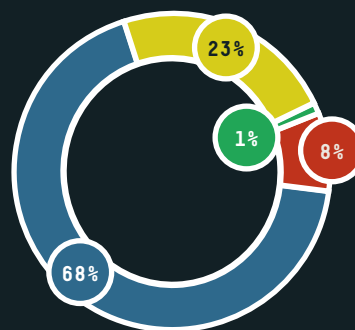
De acuerdo a la métrica MELODA¹¹ en su versión 4.13, se han cualificado los datos muestreados atendiendo a sus seis dimensiones: licencia reutilización, estándar técnico, mecanismo de acceso, modelo de datos, geolocalización de los datos y frecuencia de actualización. Los resultados obtenidos se categoriza en cuatro categorías según el grado de reusabilidad de los datos (Ilustración 20).

- ÓPTIMO
- ADECUADO
- BÁSICO
- INADECUADO

Ilustración 20:

Distribución de juegos de datos muestreados por categorías reusabilidad MELODA

∞ Fuente: Elaboración propia



Cabe destacar que sólo el 1% se encuentra en el punto óptimo de reusabilidad y que un 8% se encuentra en ámbitos inadecuados para la reutilización. La gran mayoría se encuentra en puntos iniciales (nivel básico) con muchas posibilidades de mejora.

Hay que hacer énfasis en que esta métrica mide las facilidades que se dan para el uso profesional de los datos. Estos resultados muestran que todavía hay mucho trabajo por hacer y que hay que tomar medidas para mejorar o corregir los conjuntos de datos publicados.

11. Disponible en <http://meloda.org/full-description-of-meloda/>

4.3. DIAGNÓSTICO DE LA REUTILIZACIÓN DE DATOS

4.3.1. ANÁLISIS DEL CONOCIMIENTO SOBRE LAS ENTIDADES QUE REUTILIZAN DATOS PUBLICADOS

Los resultados obtenidos en la encuesta se presentan a continuación.

Se preguntó a los encuestados si conocían a alguna entidad pública o privada (empresas, ONG, académicos, estudiantes, ciudadanos individuales) que reutilizase los datos que publican en su portal de datos abiertos.

Como se observa en la ilustración 21, el 77,8% de los encuestados afirmaron conocer a alguno de los reutilizadores; mientras que el 18,5% no los conocían y el resto no contestaron a la pregunta, lo que supone que, aproximadamente, uno de cada cinco portales no tiene información sobre sus reutilizadores.

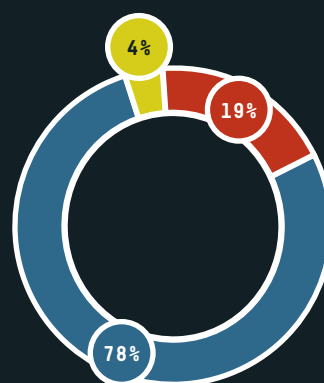
Para conocer la opinión de los encuestados sobre el grado de reutilización de los datos abiertos para cada tipo de reutilizador se preguntó en qué medida cada tipo de entidad reutilizaba los datos abiertos de su portal. La tabla 4 muestra los resultados obtenidos.



Ilustración 21:

Distribución del conocimiento de los reutilizadores de datos

∞ Fuente: Elaboración propia



Se observa que los **ciudadanos** no suelen ser reutilizadores frecuentes de estos datos, ya que sólo el 25,9% los usa frecuentemente o siempre en la opinión de los responsables de los portales.

En el caso de los **reutilizadores profesionales con ánimo de lucro** (informe-darios y desarrolladores individuales con ánimo de lucro), se obtiene que solo el 25,9% los usa frecuentemente o siempre. El cuanto al uso de los datos abiertos por **reutilizadores profesionales sin ánimo de lucro** (ONG, fundaciones, desarrolladores individuales sin ánimo de lucro y otras iniciativas sociales), se observa una frecuencia menor en el uso habitual (frecuente o siempre) de los datos, en este caso, del 22,2%.

Tabla 4:
Conocimiento del tipo de reutilizadores de datos

Fuente: Elaboración propia

ESCALA	CIUDADANOS	PROFESIONALES CON ÁNIMO DE LUCRO	PROFESIONALES SIN ÁNIMO DE LUCRO	ACADÉMICOS	PROPIA ORGANIZACIÓN	OTRAS ENTIDADES PÚBLICAS
Nunca	3,7 %	0	11,1 %	0	7,4 %	7,4 %
Casi nunca	22,2 %	18,5 %	7,4 %	14,8 %	3,7 %	14,8 %
A veces	22,2 %	18,5 %	11,1 %	25,9 %	22,2 %	22,2 %
Frecuente	11,1 %	18,5 %	11,1 %	25,9 %	22,2 %	7,4 %
Siempre	14,8 %	7,4 %	11,1 %	11,1 %	22,2 %	14,8 %
NS/NC	25,9 %	37 %	48,1 %	22,2 %	22,2 %	33,3 %
Total	100 %	100 %	100 %	100 %	100 %	100 %

Los investigadores y personal académico (incluidos estudiantes) y la propia organización son los agentes que con más frecuencia utilizan los datos abiertos. Los primeros son usuarios frecuentes o habituales en el 37% de los casos y los segundos en el 44,4%. En cambio, para otras entidades públicas, la frecuencia de uso es inferior (el 22,2% los usa frecuentemente o siempre).

También es destacable el porcentaje de casos que no saben si se están reutilizando los datos publicados en su portal. El más alto es para reutilizadores profesionales sin ánimo de lucro (48,1%), o con ánimo de lucro (37%); mientras que se tiene más información sobre el uso de datos por investigadores, académicos o por la propia organización, ya que sólo el 22,2% desconocen si estas entidades están utilizando sus datos.

De manera global, destaca que los porcentajes de desconocimiento de los reutilizadores (en general) y de los distintos tipos (en particular) es muy elevado, siendo compatibles estos resultados con una situación donde el análisis de la reutilización aún no ha sido puesto en marcha.

4.3.2. ANÁLISIS REUTILIZACIÓN DE DATOS POR SECTORES DE ACTIVIDAD

Si se analizan los sectores de actividad de los reutilizadores (Tabla 5), el primer dato que destaca es que, para la mayoría de las frecuencias de uso, en más de la mitad de los casos no disponen de esta información.

En la Tabla 5 se observa que, sobre la frecuencia de uso de los datos abiertos, se dispone de más información para los sectores de:

- Agricultura, ganadería, silvicultura y pesca
- Hostelería y turismo
- Información y comunicaciones
- Actividades profesionales, científicas y técnicas
- Administración pública y defensa
- Educación

El sector de transporte y almacenamiento y el de administración pública y defensa son los sectores que tienen mayor frecuencia de uso diario.

SECTOR	A DIARIO		FRECUENTEMENTE		ALGUNAS VECES		EXCEPCIONALMENTE		NUNCA	
	#	%	#	%	#	%	#	%	#	%
	Agricultura, ganadería, silvicultura y pesca	1	3,7			1	3,7	1	3,7	2
Industria manufacturera									1	3,7
Comercio al por mayor y al por menor			1	3,7						
Transporte y almacenamiento	3	11,1								
Hostelería y turismo	1	3,7	2	7,4	3	11,1	1	3,7	1	3,7
Información y comunicaciones	1	3,7	3	11,1			2	7,4		
Actividades financieras y de seguros					1	3,7				
Actividades inmobiliarias	1	3,7			2	7,4				
Actividades profesionales, científicas y técnicas			2	7,4	2	7,4	2	7,4		
Actividades administrativas y servicios auxiliares			1	3,7						
Administración Pública y defensa	3	11,1	3	11,1	3	11,1	2	7,4		
Educación			1	3,7	2	7,4	2	7,4		
Actividades sanitarias y de servicios sociales							1	3,7		
Actividades artísticas, recreativas y de entretenimiento					1	3,7				
Otros servicios	1	3,7			1	3,7				
Actividades de organizaciones y organismos extraterritoriales					1	3,7	1	3,7		
NS/NC	16	59,3	14	51,9	10	37,0	15	55,6	23	85,2

Tabla 5:
Reutilizadores por sector de actividad

∞ Fuente: Elaboración propia

Los sectores de información y comunicaciones y de administración pública y defensa son los que destacan en el uso frecuente de datos respecto de los demás sectores. Aun así, su frecuencia es baja (11,1%, en cada caso).

4.3.3.

ANÁLISIS REUTILIZACIÓN DE DATOS POR ÁMBITO TERRITORIAL

Para conocer los ámbitos de actuación, se preguntó a los portales por aquellos en los que trabajaban los reutilizadores de sus datos (Tabla 6).

ESCALA	LOCAL		AUTONÓMICO		NACIONAL		EUROPEO		OTROS	
Nunca (1)	1	3,7 %	0	0 %	2	7,4 %	3	11,1 %	2	7,4 %
Casi nunca (2)	3	11,1 %	3	11,1 %	2	7,4 %	3	11,1 %	2	7,4 %
A veces (3)	5	18,5 %	6	22,2 %	4	14,8 %	2	7,4 %	0	0 %
Frecuentemente (4)	3	11,1 %	2	7,4 %	4	14,8 %	1	3,7 %	1	3,7 %
Siempre (5)	4	14,8 %	3	11,1 %	0	0 %	1	3,7 %	2	7,4 %
NS/NC	11	40,7 %	13	48,1 %	15	55,5 %	17	62,9 %	20	74,1 %
Total	27	100 %	27	100 %	27	100 %	27	100 %	27	100 %

Tabla 6:

Ámbito de actuación de los reutilizadores de datos abiertos

∞ Fuente: Elaboración propia

Al igual que ocurría con los sectores, se observa que, en más de la mitad de los casos, en cada uno de los ámbitos analizados, salvo el caso del ámbito local, 40,7%, no se dispone de información sobre el uso de los datos.

Si se analizan para cada ámbito los casos en los que los usos son “frecuente” o “siempre”, se observa que el ámbito local tiene un porcentaje del 25,9%; mientras que en el autonómico (18,5%), el nacional (14,8%) y el europeo (7,4%), las cifras son menores.

También es destacable el resultado obtenido para la frecuencia “nunca”. En este caso, es el ámbito europeo el que tiene una cifra mayor (11,1%), siendo la del nacional el 7,4% y la del local el 3,7%.

4.3.4.

ANÁLISIS DE LOS TIPOS DE INNOVACIÓN POR LA REUTILIZACIÓN DE DATOS ABIERTOS

Para conocer los tipos de innovación que se pueden lograr con el uso de datos abiertos, se preguntó a los portales en qué medida se utilizaban los tipos de innovación que aparecen en la tabla 7.

Por el tipo de innovación, los resultados obtenidos muestran que en el 44,4% de los casos, los encuestados consideran que con el uso de los datos abiertos de su portal nunca, casi nunca o a veces se realizan innovaciones de producto. Los mismos resultados se obtienen para la innovación de proceso. Estas cifras ponen de manifiesto la falta de información que hay sobre el tipo de actividades que los reutilizadores realizan con los datos de los portales de datos.

Tabla 7:
Tipos de innovación por la reutilización de datos abiertos

Fuente: Elaboración propia

ESCALA	INNOVACIÓN DE PRODUCTO		INNOVACIÓN DE PROCESO	
Nunca	2	7,4 %	2	7,4 %
Casi nunca	7	25,9 %	6	22,2 %
A veces	2	7,4 %	3	11,1 %
Frecuentemente	4	14,8 %	3	11,1 %
Siempre	0	0 %	1	3,7 %
NS/NC	12	44,4 %	12	44,4 %
Total	27	100 %	27	100 %

4.3.5. ANÁLISIS DE LA DISPONIBILIDAD DE REGISTRO DE LOS ACCESOS

La tabla 8 muestra la información disponible según los responsables de los portales de acuerdo a los registros de acceso.

Tabla 8:
Disponibilidad de los registros de acceso a los datos

Fuente: Elaboración propia

	PORCENTAJE
No	25,9
Sí	59,3
NS/NC	14,8
Total	100,0

Los registros de los accesos a los datos son un elemento imprescindible para poder gestionar la demanda y poder promoverla. Sería la herramienta más básica de gestión del conocimiento de la demanda de los distintos datos publicados.

4.3.6. ANÁLISIS DE LAS ACTIVIDADES DE PROMOCIÓN DEL USO DE DATOS ABIERTOS

Para conocer las actividades de promoción del uso de datos abiertos se preguntó a los portales en qué medida utilizaban algunos tipos de actividades de promoción. En la tabla 9 se muestran los resultados obtenidos.

Respecto a las actividades promoción del uso de datos abiertos, en ninguno de los casos se utiliza siempre una actividad. Sin embargo, sí que hay actividades que se utilizan de forma frecuente. Las reuniones con reutilizadores son las más habituales (26%), seguidas por los actos externos (organizados por otras entidades) de presentación de los datos disponibles (22,2%), los actos propios (18,5%). Por otra parte, los concursos de aplicaciones (11%) son actividades que realizan prácticamente uno de cada diez portales.

ESCALA	CONCURSOS DE APLICACIÓN		ACTOS PROPIOS DE PRESENTACIÓN		ACTOS EXTERNOS DE PRESENTACIÓN		REUNIONES CON REUTILIZADORES		OTROS	
Nunca	11	41%	2	7%	2	7%	7	26%	5	19%
Casi nunca	4	15%	8	30%	8	30%	7	26%	0	0%
A veces	3	11%	8	30%	6	22%	1	4%	1	4%
Frecuentemente	3	11%	5	18%	6	22%	7	26%	2	7%
Siempre	0	0%	0	0%	0	0%	0	0%	0	0%
NS/NC	6	22%	4	15%	5	19%	5	18%	19	70%
Total	27	100%	27	100%	27	100%	27	100%	27	100%

Tabla 9:
Actividades de promoción del uso de datos abiertos

∞ Fuente: Elaboración propia

También se obtienen casos en los que los encuestados consideran que nunca se utilizan estas actividades, destacándose las cifras obtenidas para los concursos de aplicaciones (40,7%) y las reuniones con reutilizadores (25,9%).

Ninguna de las actividades muestreadas tiene como respuesta la opción *siempre*. Esto refleja que la promoción de los datos no es una actividad sistemática de los responsables de los portales..

4.4. DIAGNÓSTICO DE LOS SERVICIOS GENERADOS

Para identificar los servicios generados, por un lado, se preguntó a los responsables de los portales por los usos más comunes de los datos del portal (anexo 62) y por otro, los juegos de datos más utilizados (anexo 64).

4.4.1. ANÁLISIS DE LAS TEMÁTICAS DE LOS SERVICIOS

Para obtener información más detallada sobre los servicios generados se realizó un análisis de los mismos a partir de la información obtenida en los juegos de datos de los portales. Por temáticas de los servicios, las categorías más frecuentes son las de temática Otras (24%)¹², seguida de Transporte, (22%) y Geográfica (16%).

4.4.2. ANÁLISIS DE SOSTENIBILIDAD Y MODELOS DE NEGOCIO DE LOS SERVICIOS

También se ha analizado si los servicios desarrollados poseían algún mecanismo de sostenibilidad (patrocinio de una institución o mecanismos de generación de ingresos, bien directos o indirectos (pagos *freemium* o publicidad). El 52% de las aplicaciones declaró no tener ningún mecanismo de sostenibilidad.

Además, de éstas que tienen mecanismos de sostenibilidad, sólo la mitad, aproximadamente, presentaban un modelo de negocio, es decir, poco más de un 25% del total.

12. Entre otras temáticas incluye turismo.

4.4.3. ANÁLISIS DE LOS AUTORES DE LOS SERVICIOS

De igual manera, el muestreo de aplicaciones indica que lo más frecuente es que los generadores de los servicios sean organizaciones públicas (43%, tanto la que publica los datos como otras relacionadas), mientras que los reutilizadores profesionales dan cuenta del 30% y los ciudadanos y desarrolladores individuales del 18% (Ilustración 22).

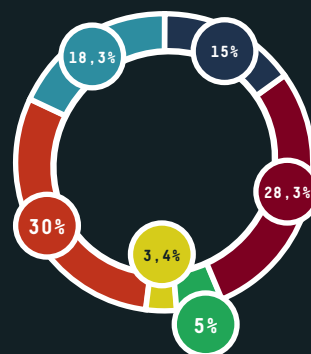
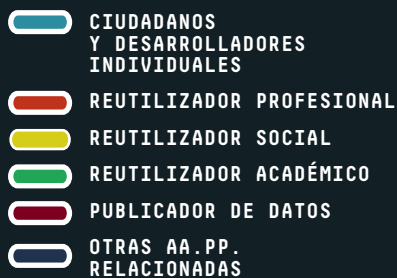


Ilustración 22: **Distribución de autores de los servicios**

∞ Fuente: Elaboración propia

4.4.4. ANÁLISIS DE OTRAS CARACTERÍSTICAS DE LOS SERVICIOS

Existe una muy elevada tendencia a que los servicios prestados estén geolocalizados (72%); sin embargo, sólo el 35 % son servicios con información en tiempo real. Esto contrasta de forma importante con las informaciones publicadas, que en más de la mitad de los casos no contenían información geográfica.

También contrasta de forma destacada con el hecho de que la información se actualice, en general, de forma muy lenta (con un tiempo mayor de una semana) y, sin embargo, uno de cada tres servicios lo sea en tiempo real.

4.4.5. ANÁLISIS DE LA CREACIÓN DE VALOR DE LOS SERVICIOS

Para conocer la creación de valor, se preguntó a los portales en qué medida la reutilización de datos ha creado valor en los siguientes ámbitos para las entidades que utilizan los datos o para sus clientes / usuarios (Tabla 10).

Uno de los aspectos más relevantes cuando se analiza la reutilización de los datos abiertos es la creación de valor que se puede lograr. Sin embargo, realizar un análisis de la creación de valor no es fácil y en muchos casos es una tarea pendiente.

ESCALA	CLIENTES SATISFECHOS	CIUDADANOS SATISFECHOS	MEJORA MEDIOAMBIENTAL	MEJORA INFRAESTRUCTURAS	CREACIÓN DE EMPRESAS	MEJORA SERVICIOS AA. PP.	OTROS
Nunca	11,1	7,4	14,8	7,4	14,8	7,4	11,1
Casi nunca	3,7	7,4	0	0	0	3,7	0
A veces	3,7	7,4	11,1	7,4	18,5	33,3	3,7
Frecuente	18,5	25,9	18,5	14,8	0	11,1	0
Siempre	11,1	11,1	0	14,8	0	22,2	0
NS/NC	51,8	40,7	55,5	55,5	66,6	22,2	85,2
Total	100	100	100	100	100	100	100

Tabla 10:
Creación de valor por la reutilización de los datos

∞ Fuente: Elaboración propia

En la encuesta realizada destacan los porcentajes obtenidos para la ausencia de respuesta o para la opción “no sabe”: 51,8% para el criterio de satisfacción del cliente, 40,7% para el de satisfacción del ciudadano, 55,5% para la mejora medioambiental, 55,5% para la mejora de infraestructuras, 66,6% para la creación de empresas y 85,2% para otros.

El 37% de los responsables de portales considera que, frecuentemente o siempre, los ciudadanos se sienten más satisfechos con la publicación de datos. La mejora de servicios de la Administración Pública (AA. PP.) es el criterio para el que se dispone de más información, donde se observa que, en el 33,3% de los casos, se logran de forma frecuente o siempre mejoras en las AA. PP. Sin embargo, la reutilización de los datos abiertos permite de forma muy poco frecuente la creación de empresas (14,8%) o a veces (18,5%). Siendo éste el mecanismo de creación de valor menos probable de los encuestados en la opinión de los responsables de los portales de datos.

Respecto a los criterios de satisfacción del cliente y del ciudadano, aunque son dos aspectos fundamentales para medir la creación de valor, no presentan unos porcentajes muy altos en las categorías frecuentemente (para los clientes, 18,5% y para los ciudadanos, 25,9%) y siempre (tanto para clientes como para ciudadanos es del 11,1%).

En el caso de la mejora medioambiental y de infraestructuras, se obtiene que el 18,5% de los encuestados consideran que la reutilización de los datos permite frecuentemente la mejora medioambiental y el 14,8% que permite frecuentemente la mejora de infraestructuras, mientras que el 14,8%, afirma que siempre hay una mejora de estas últimas.

ESTIMACIÓN CUALITATIVA DE SERVICIOS INNOVADORES

5.1.

TIPOS DE MODELOS DE NEGOCIO IDENTIFICADOS

En nuestro caso, a partir del estudio realizado sobre los servicios muestreados, se han identificados cuatro modelos de negocio:

- Basado en datos de los usuarios
- Financiación por publicidad
- Servicios Freemium
- Basados en recomendaciones contextuales

5.1.1.

DATOS DE LOS USUARIOS

El primer modelo de negocio es la gestión de los datos de los usuarios, por ejemplo, su movilidad, gastos, ubicación de los usuarios para poder conectarlos a nivel global con grandes operadores, telecomunicación, automoción, etc. Este tipo de alianzas para compartir datos proporcionaría la financiación suficiente para el sostenimiento del servicio.

Bien es cierto que este tipo de modelos de negocio, salvo excepciones, requieren de dimensiones globales (aunque sea entre colectivos reducidos) para ser viables. Un ejemplo ya mencionado es la aplicación Moovit (<https://moovitapp.com/>), presente en más de 60 países y que comenzó por España su internacionalización gracias a la disponibilidad de los datos de transporte de la Empresa Municipal de Transporte de Madrid.

5.1.2.

ANUNCIOS

Posiblemente el más conocido, y también el más incierto, dado que depende de los precios de pago de la publicidad mostrada y de la capacidad tractora del servicio para que la publicidad sea personalizada y efectiva. Este modelo de negocio sería equivalente al identificado en Berends *et al.* (2017), como proveedor de datos (data supplier).

Un ejemplo sería la aplicación Sevici, que además del acceso a los datos del servicio de bicicletas de la ciudad de Sevilla contiene anuncios que ayudan a la financiación del desarrollador individual (Michael Gauzins, basado en Francia). Además, la aplicación tiene un coste de 2.99€ en el Android Store, actualmente tiene entre 5.000 y 10.000 descargas y una valoración media por sus usuarios de 4,6 sobre 5.

5.1.3.

FREEMIUM

En este caso se publican algunas características de interés general de forma gratuita y otras de carácter más profesional bajo cuotas de pago. Aquí el desafío es la tasa de conversión de un tipo de cliente (free) en el de pago (premium).

Un ejemplo de este modelo de empresa es Eixos (<http://eixos.cat>), la cual recopila datos públicos y genera sus propios datos para proporcionar servicios de marketing geolocalizado sobre puntos de venta. Algunos de sus datos son de libre acceso y otros requieren la contratación de servicios.

5.1.4.

RECOMENDACIONES CONTEXTUALES

Dado que la ubicación del usuario es mayormente conocida para el servicio, éste puede lanzar ofertas personalizadas al momento, lugar y gustos del usuario. Se puede extraer valor de los anunciados en la aplicación. Nuevamente la necesidad de cobertura intensa y global es el desafío de alcanzar la sostenibilidad.

Un ejemplo sería bikecitizens, una aplicación que, alrededor del mundo del desplazamiento en bicicleta, proporciona toda una serie de posibilidades de promoción localizada para compañías, co-branding, etc. Se puede obtener más información en su portal inicial <https://www.bikecitizens.net> y en su opción para negocios <https://www.bikecitizens.net/partner/companies/>

5.2.

OTROS MECANISMOS DE SOSTENIBILIDAD DEL USO DE DATOS ABIERTOS

Algunas instituciones generan servicios y aplicaciones que reutilizan datos abiertos, no para proporcionar servicios a sus ciudadanos o clientes sino con propósitos, como por ejemplo, de marketing, de transparencia, publicidad, promoción turística, etc.

De esta manera, en el análisis de servicios, se han identificado un número significativo de servicios (25% del total) que, sin tener un modelo de negocio que pudiera sostenerlo, es decir, sin tener mecanismos que generen ingresos para el servicio, sí que tienen detrás una entidad que soporta sus gastos de puesta en marcha y explotación.

Hay ejemplos donde se entiende que el servicio reutilizador de datos (P. ej. publicar agendas de actos festivos) se utiliza como herramienta de marketing corporativo. O directamente las propias AA. PP. asumen el coste del mantenimiento del servicio reutilizador de datos como vía alternativa de comunicación con el ciudadano.

5.3. ANÁLISIS DE LOS SERVICIOS POR EL MODELO DE NEGOCIO

Se observa un desajuste entre las características de los datos que se publican y las de los servicios que se generan; especialmente en aspectos como la frecuencia de actualización de los datos y en sus características de geolocalización. Detalles complementarios se incluyen en el punto 3.4.4.

De hecho, el 87% de los servicios con un modelo de negocio están geolocalizados, por el contrario, sólo se da en el 67% de los servicios cuando no existe un modelo de negocio. Pero donde se encuentra la diferencia más grande es en la variable de tiempo real. Dos de cada tres servicios desarrollados con modelo de negocio son en tiempo real. Este porcentaje cae a uno de cada cuatro cuando se carece de modelo de negocio

5.4. COMPARATIVA DE SERVICIOS GENERADOS VS DATOS DISPONIBLES

De las aplicaciones con modelo de negocio, destacan las de transporte (47% del total) y las de temática meteorológica (27%). Esto contrasta con los datos que se publican en estas áreas y que representan sólo el 3,9 % para el transporte, y el 6,8 % para la información medioambiental, donde se enmarca la información meteorológica.

Los resultados de la encuesta a los responsables de datos muestran una falta de datos o de conocimiento sobre las innovaciones generadas y por tanto no es posible extraer conclusiones sobre los mismos.

Informaciones minoritarias en porcentaje de datos publicados como transporte, 4% y meteorología, 7% son la base de la mayoría de los servicios generados, 47% y 27% respectivamente.

Este análisis considera el entorno interno, incluyendo a los gestores de los portales y los reutilizadores de todo tipo de datos, y el entorno externo (Tabla 11).

DEBILIDADES	AMENAZAS
<p>Falta de utilización de herramientas de gestión de portales de datos (DMS) en portales menos maduros (3.1.2)</p> <p>Información débilmente geolocalizada y con bajas frecuencias de actualización (3.2.6)</p> <p>Conocimiento limitado de los reutilizadores (3.3.1)</p> <p>Registro del acceso y conocimiento del uso de los datos publicados (3.3.5)</p> <p>Servicios con modelos de negocio sostenible (3.4.2)</p> <p>Desconocimiento por los gestores de los portales de la creación de valor (3.4.5)</p> <p>Falta de sistemática en la promoción del uso de los datos publicados (4.3.6)</p> <p>Desalineación entre las características y temáticas de los datos publicados y los servicios generados (4.4)</p>	<p>El volumen de datos, y su uso para la asignación de recursos</p> <p>Objetivos desalineados entre la organización y los gestores de los portales de datos</p> <p>Dificultad para demostrar la innovación generada con la reutilización de datos</p>
FORTALEZAS	OPORTUNIDADES
<p>Madurez del sector de datos abiertos en España (1.1)</p> <p>La mayoría de los juegos de datos se encuentran en portales con grado de madurez más elevado (3.2.1)</p> <p>Alta difusión de los datos por los reutilizadores académicos (3.3.1)</p>	<p>Mayor comunicación con los reutilizadores</p> <p>Coordinación de las actividades de publicación y promoción del uso de todos los portales a nivel nacional</p> <p>Desarrollo de servicios basados en datos orientados a la creación de empresas</p> <p>Desarrollo de más servicios basados en los datos orientados a atender las expectativas ciudadanas</p>

Tabla 11:

Análisis DAFO

∞ Fuente: Elaboración propia

CONCLUSIONES

La sociedad actual puede ser calificada como la sociedad de los datos. Además, se prevé que ésta dinámica se acelere en los próximos años.

España ha conseguido situarse como país referencia en materia de publicación de datos desde el ámbito público, a nivel europeo y mundial y, por tanto, se sitúa en una posición de partida idónea. Existe un número de fuentes de datos considerable y, se comienza entender que la publicación de datos es una tarea de servicio público, e incluso, en algunos casos, se establecen ordenanzas y normativas que así lo reflejan.

Además, el gran número de servicios basados en datos abiertos desarrollados por AA. PP., reflejan que éstas han dado apoyo a las innovaciones basadas en datos.

7.1. PRINCIPALES BARRERAS

Las principales barreras identificadas a partir de los análisis realizados son:

- **Una muy pequeña proporción de los datos publicados son utilizados para la generación de servicios innovadores, tal y como se vio en el punto 4.4.**
- **La gran mayoría de los datos publicados no tienen las características necesarias para generar servicios innovadores.**

7.2. OTRAS BARRERAS

Así mismo, se han detectado otros tipos de barreras que pueden ser clasificadas en dos grupos. El primero, denominado barreras estratégicas, se refiere a la organización de los publicadores de datos para generar el entorno donde puedan surgir los servicios. El segundo grupo abarca aspectos técnicos de la reutilización de datos para la generación de negocios.

7.2.1.

**BARRERAS
ESTRATÉGICAS**

La publicación sistemática de datos es una decisión estratégica de las organizaciones públicas que ha sido progresivamente adoptada, bien por normativas o bien por planes estratégicos. Esa estrategia debe reflejarse en cambios estructurales, que, según los datos analizados, se encuentran todavía pendientes.

Una gobernanza interna de datos, permite *controlar y gestionar* cómo se mueven los datos por el interior de la organización y, por tanto, habilita la publicación sistemática de datos con las características idóneas de temática, frecuencia de actualización y referencia geográfica.

Para que los datos publicados generen servicios han de ponerse en marcha mecanismos de **conexión con los reutilizadores**. El análisis realizado refleja una importante carencia de datos sobre los usos industriales y sobre los reutilizadores de los datos publicados, así como una falta de sistemática en la relación con los mismos. Tanto, que incluso los propios responsables de los portales califican como último mecanismo de creación de valor la creación de empresas.

7.2.2.

BARRERAS TÉCNICAS

Siendo, posiblemente, las más sencillas de resolver, todavía están pendientes en la mayoría de los casos. Específicamente, el poco uso de **herramientas adecuadas (DMS)** para la publicación de datos, limita la automatización del acceso, la publicación sistemática y el aviso de las novedades. Incluso limita la comprensión de la información al gran público por las deficiencias en la visualización de los datos publicados.

También, aunque puntualmente, se encuentran barreras en los estándares de publicación y en mecanismos de acceso no adecuados para la reutilización de los datos y la generación de servicios.

7.3.

**MEDIDAS PARA
SOBREPASAR
LAS BARRERAS**

La financiación de las iniciativas de datos sería mucho más sencilla en el caso de que se demuestre el impacto generado por los servicios basados en ellos.

Pese a que el análisis de este sector es ejemplar a nivel mundial (Red.es, 2011, 2012, 2015, 2017), los análisis de su impacto están limitados por la falta de un modelo al uso (Abella *et al.*, 2017b). Algunas de las propuestas son:

- **Plan de digitalización pública y gobernanza de datos.** Permitiría que los datos fluyeran tanto interna como externamente de la forma necesaria para la generación de servicios.
- **Actualización de las infraestructuras de publicación de datos.** La no disponibilidad de herramientas apropiadas está limitando el descubrimiento de los datos publicados y el desarrollo de servicios.

- **Estrategia de generación de servicios basados en datos.** La publicación de datos no es un ejercicio de transparencia sino la siembra de unos recursos que deben fructificar en otras entidades (reutilizadores). Ha de cuidarse el campo donde ésta reutilización se produce, prestando atención y analizando los usos que se realizan y proporcionando soporte a los mismos.
- **Coordinación de esfuerzos (Asociación / grupo de publicadores).** El gran número de entidades publicando datos contrastan con la falta de organizaciones o grupos tractores dedicados a coordinar esta publicación. La dispersión de formatos y tecnologías es un buen ejemplo de ello. Esta asociación/grupo permitiría reducir esfuerzos comunes y facilitar la generación de servicios innovadores basados en los datos.

7.3.1.

ASOCIACIÓN DE PUBLICADORES DE DATOS

La publicación de datos y su reutilización es un fenómeno creciente tanto en su complejidad como en el número de actores implicados. Actualmente, se cifra en más de 20.000 los juegos de datos disponibles en España.

Los datos publicados han de ser reutilizados para que tengan impacto sobre la sociedad, favoreciendo la actividad económica, la innovación basada en productos o servicios o las nuevas aplicaciones sociales.

De acuerdo a a datainception.io (Julio de 2017) en Europa hay actualmente más de 2.700 entidades publicando datos y solo en el portal paneuropeo de datos se recopilan más de 760.000 juegos de datos (Julio de 2017). Hay que tener en cuenta que dicho portal comenzó a recopilar datos desde noviembre de 2015 y en el momento de su lanzamiento el número de datos recopilados únicamente superaba los 240.000.

Necesidades identificadas en España por informes de RED.ES

Las necesidades identificadas en el sector nacional por el informe del sector infomediario elaborado por la entidad Red.es (2017) del Ministerio de Energía, Turismo y Agenda digital son:

- **R1. Necesidad de homogeneizar la información en las CC. AA. y las EE. LL. Y, en general, en toda la Administración.**
- **R2. Avance de la proactividad en la gestión de la información por las Administraciones.**
- **R3. Fomentar el desarrollo de herramientas TIC para el tratamiento de los datos públicos.**
- **R4. Mejorar la disponibilidad y calidad de los datos públicos.**
- **R5. Un listado de ámbitos en los que el sector requiere mejoras en la disponibilidad o acceso a los datos.**

Necesidades identificadas en España por informes de ASEDIE

Los problemas identificados por los reutilizadores (ASEDIE, 2017) de España son los siguientes.

- **A1. Elaborar, revisar, aclarar y/o priorizar las distintas normativas existentes.**
- **A2. Bases de datos por abrir (o necesidad de mayor detalle).**
- **A3. Acceso transfronterizo a los datos.**
- **A4. Interconexión de registros.**
- **A5. Libre circulación de datos a nivel europeo.**

Propuesta de una asociación de publicadores de datos

Una asociación de publicadores de datos podría:

- 1. Ayudar a homogeneizar las informaciones de las distintas fuentes y reducir los esfuerzos de publicación mediante modelos de datos compartidos que faciliten la interconexión de registros (R1 y A4).**
- 2. Facilitar la extensión de las bases de datos publicadas (R2, R4, R5 y A2).**
- 3. Fomentar interpretaciones comunes de las normativas existentes o promover la generación de nuevas (A1).**
- 4. Fomentar la reutilización conjunta de sus datos y así el impacto en la sociedad.**

Con un carácter europeo, además, podría abordar las problemáticas A3, A4 y A5.

También debería incluir algunos mecanismos para poder adaptarla a la velocidad con la que se está generando nuevos datos, ya que los mecanismos existentes (entidades de normalización no pueden responder en los tiempos exigidos por este nuevo sector). La carencia de una entidad que asuma estas funciones puede provocar que los esfuerzos globales de esos más de 700 publicadores de datos en la UE (según datainception.io) se vean disminuidos o incluso neutralizados ante el coste del acceso efectivo y automatizado a los datos y las transformaciones necesarias para la generación de nuevos modelos de negocio.

7.4. FUTURAS LÍNEAS DE TRABAJO

7.4.1. ESTUDIO DE LA INNOVACIÓN EMPRESARIAL A PARTIR DE LA REUTILIZACIÓN DE DATOS ABIERTOS

Estudio de las innovaciones de producto, proceso, organizativa y comercial basadas en datos. Para ello, dado que los reutilizadores de datos no pueden ser localizados desde los portales de datos, existen otras fuentes alternativas como son los datos en poder de Red.es para la elaboración de informes del sector, o los disponibles en otras organizaciones como ODINE, FINODEX, que son incubadoras de negocios innovadores basados

en datos. Una colaboración con estas últimas, permitiría un análisis europeo con una muestra significativa en el tramo empresarial.

7.4.2. MÉTRICAS DE REUTILIZACIÓN DE DATOS

Las métricas de reutilización de datos, si bien proporcionan resultados interesantes han de ser automatizables y sencillas de aplicar para la mejora de la publicación de datos. Actualmente, todavía tienen una importante componente manual que limita su uso. Sería posible incorporarlas en las principales herramientas de publicación de datos (P. ej. CKAN) en forma de *plugin*.

7.4.3. ESTUDIO DE LAS CONDICIONES DE GENERACIÓN DE NUEVOS NEGOCIOS DESDE LAS FUENTES DE DATOS

Si bien se han detectado servicios generados basados en los datos publicados, falta un elemento que determine cuáles son las condiciones suficientes y cuáles son los incentivos adecuados para la generación de servicios desde los datos publicados.

7.4.4. ACTUALIZACIÓN CONTINUA DE LOS DATOS

Durante la ejecución del proyecto se han detectado la actualización de diversos portales de datos (Junta Andalucía, Ayuntamiento de Barcelona), lo que por una parte identifica un sector en evolución pero que requeriría de un seguimiento continuo de su evolución. Esto podría coordinarse con otros observatorios de la realidad de los datos de España y la UE, por ejemplo, el observatorio para la sociedad de la información de Red.es y el portal paneuropeo de datos con quien se ha tenido ya contacto durante la realización del informe.

BIBLIOGRAFÍA

- Abella, A. (2017). Modelling of the social and economic impact of open data in smart cities' ecosystems. Tesis doctoral. Escuela Internacional de Doctorado. Universidad Rey Juan Carlos.
- 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.
<http://dx.doi.org/10.3145/epi.2014.nov.04>
- Abella, A.; Ortiz-de-Urbina-Criado; M., De-Pablos-Heredero, C. (2015). The ecosystem of services around smart cities: An exploratory analysis. Procedia Computer Science, 64, 1075-1080
<http://dx.doi.org/10.1016/j.procs.2015.08.554>
- Abella, A.; Ortiz de Urbina Criado, M.; De Pablos Heredero, C. (2017a). Social and Economic Impact of Open Data Reuse in the International Context. Comunicación oral presentada en la XIX Reunión de Economía Mundial / World Economy Meeting XIX, Universidad de Huelva, 10-12 de mayo de 2017.
- Abella, A., Ortiz-de-Urbina-Criado, M.; De-Pablos-Heredero, C. (2017b). A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems. Cities, 64, 47-53.
- ASEDIE. (2017). Informe del sector infomediario (V Edición). Disponible en **<http://www.asedie.es/assets/informe-sector-infomediario-2017.pdf>**
- Berends, J.; Carrara, W.; Engbers, W.; Vollers, H. (2017). Reusing open data. An study on companies transforming Open Data into economic & societal value. European Union.
- Carrara, W.; San Chan, W.; Fischer, S.; Van Steenberghe, E. (2015). Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources. European commission. Directorate General for Communications Networks, Content and Technology. Unit G.3 Value Data Chain
- Carrara, W.; Nieuwenhuis, M.; Vollers, H. (2016). Open data in Europe 2016. Insights into the European state of play. European Commission. Directorate General for Communications Networks, Content and Technology.
- Dekkers, M., Polman, F., Te Velde, R., De Vries, M. MEPSIR (Measuring European Public Sector Information Resources) (2006), "Final report of study on exploitation of public sector information – benchmarking of EU framework conditions", Executive summary and Final report Part 1 and Part 2
- European project LAPSI II. (2015). Legal aspects of public sector information (LAPSI) thematic network outputs. Entregables disponibles en: **<https://ec.europa.eu/digital-single-market/en/news/legal-aspects-public-sector-information-lapsi-thematic-network-outputs>**
- European Union (2017). Application Profile for data portals in Europe Version 1.1. Disponible en: **https://joinup.ec.europa.eu/asset/dcat_application_profile/description**

- Ferrer-Sapena, A.; Peset, F.; Aleixandre-Benavent, R. (2011). "Acceso a los datos públicos y su reutilización: open data y open government". *El profesional de la información*, 20/3, 260-269.
<http://dx.doi.org/10.3145/epi.2011.may.03>
- 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.; Bjørn-Andersen, N. (2013). Generating value from open government data. *Thirty Fourth International Conference on Information Systems*, Milan 2013. 1-19
- Jetzek, T., Avital, M., Bjorn-Andersen, N. (2014a). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research* 9/2, 100-120.
- Jetzek, T., Avital, M.; Bjørn-Andersen, N. (2014b). Generating sustainable value from open data in a sharing society. In *International Working Conference on Transfer and Diffusion of IT* (pp. 62-82). Springer Berlin Heidelberg.
- Lee, G.; Kwak, Y.H. (2012). "An Open Government Maturity Model for social media-based public engagement". *Government Information Quarterly*, 29, 492-503. **<http://dx.doi.org/10.1016/j.giq.2012.06.001>**
- Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. Disponible en:
<https://www.boe.es/boe/dias/2015/07/10/pdfs/BOE-A-2015-7731.pdf>
- Marcos-Martín, C.; Soriano-Maldonado, S.L. (2011). Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta. *El profesional de la información*, 20/3, 291-297.
<http://dx.doi.org/10.3145/epi.2011.may.07>
- Red.es. (2011). Characterization Study of the Infomediary Sector 2011. Disponible en **http://datos.gob.es/datos/sites/default/files/files/aporta_infomediary%20sector_2011.pdf**
- Red.es. (2012). Characterization Study of the Infomediary Sector 2012. (2012). Disponible en **http://www.ontsi.red.es/ontsi/sites/default/files/121001_red_007_final_report_2012_edition_vf_en_1.pdf**
- Red.es. (2015). Estudio de Caracterización del Sector Infomediario en España 2014 (Reutilización de Información del Sector Público). Disponible en **http://www.ontsi.red.es/ontsi/sites/default/files/estudio_de_caracterizacion_del_sector_infomediario_en_espana_2014_parte_i_publica.pdf**
- Red.es (2017). "Caracterización del Sector Infomediario en España 2016. Disponible en **http://www.ontsi.red.es/ontsi/sites/ontsi/files/Estudio%20de%20Caracterizaci%C3%B3n%20del%20Sector%20Infomediario%202016_o.pdf**
- Resolución de 19 de febrero de 2013, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información. Disponible en **<https://www.boe.es/boe/dias/2013/03/04/pdfs/BOE-A-2013-2380.pdf>**

9.1. RELACIÓN DE PORTALES DE DATOS MUESTREADOS

Relación de portales analizados entre los meses de marzo y abril de 2017

- <http://datos.zaragoza.es>
- <https://opendata.aemet.es>
- <http://opendata.badalona.cat/badalona/ca>
- <http://pamplona.es/verpagina.asp?idpag=1519&idioma=1>
- <http://www.bcn.cat/opendata/>
- <http://datos.gob.es>
- <http://www.opendatacanarias.es/>
- <http://datosabiertos.malaga.eu/>
- <http://datos.santander.es/>
- <http://opendatamarbella.es/>
- <http://opendata.caceres.es/>
- <http://gobiernoabierto.valencia.es>
- <http://opendata.ugr.es/>
- <http://concellolugo.esrism.opendata.arcgis.com>
- <http://dadesobertes.situam.opendata.arcgis.com>
- <http://www.juntadeandalucia.es/datosabiertos/portal.html>
- <http://opendata.euskadi.net>
- <https://sede.puertodemelilla.gob.es/index.php/datos-abiertos.html>
- <http://datos.lorca.es/>
- <http://gobiernoabierto.navarra.es/es/open-data>
- <https://opendata.rubi.cat/en/>
- <https://opendata.granollers.cat/>
- <https://gavaobert.gavaciutat.cat/en/>
- <http://opendata.terrassa.cat/>
- <http://www.datosabiertos.jcyl.es/>
- <http://www.caib.es/caibdatafront/>
- <http://opendata.elprat.cat/elprat/en>

- <http://opendata.aragon.es>
- <http://abertos.xunta.es/portada>
- <http://www.ine.es/>
- <http://www.malaga.es/gobiernoabierto/>
- <http://opendata.jccm.es/>
- <http://dadesobertes.gencat.cat>
- <http://data.upf.edu/es/main>
- <http://datos.gijon.es>
- <http://www.gipuzkoairekia.eus>
- <http://www.bilbao.net/opendata/>
- <http://datos.alcobendas.org>
- <http://datos.madrid.es>
- <http://opendata.santfeliu.cat/santfeliu/en>
- <http://opendata.tarragona.cat/tarragona/en>
- <http://opendata.viladecans.cat/viladecans/en>
- <http://dadesobertes.diba.cat>
- https://www.elpuertodesantamaria.es/index.php?men_id=436&art_id=9262&PRY=
- <http://datos.crtm.es/>
- <http://www.asturias.es/portal/site/webasturias/menuitem.a76385ecc651687bd9db8433f2300030/?vgnextoid=698b5docfd524410VgnVCM100000ce212boaRCRD>
- <http://opendata.manresa.cat/>
- <http://opendata.emtmadrid.es>
- <https://seu-e.cat/es/web/santboidellucanes/dades-obertes>
- <http://opendata.sabadell.cat/ca/inici>
- <http://datos.diputacionalicante.es>
- <http://www.opendatalapalma.es/>
- <http://datosabiertos.dphuesca.es>
- <https://datosabiertos.dipucadiz.es>
- <http://datosabiertos.dipc.as.es/>
- https://sede.dipucuenca.es/sede/portal-de-datos-abiertos-de-la-diputacion-de-cuenca?p_p_auth=wBQUfmCE&p_p_id=49&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&_49_struts_action=%2Fmy_sites%2Fview&_49_groupId=10182&_49_privateLayout=false
- <http://www.paeria.cat/opendata/ca/index.asp>

- <http://www.porttarragona.cat/es/gestio-portuaria/dades-obertes.html>
- <https://opendata.terrassa.cat/>
- <http://www.amb.cat/es/web/area-metropolitana/dades-obertes>
- <https://dadesobertes.grupsagessa.com/>
- <http://transparencia.ajsolsona.cat/dades-obertes>
- <http://datos.arona.org/>
- <http://datos.aytocamargo.es/>
- <https://smart.coruna.es/OpenData/COR/index.html>
- <http://opendata.uji.es/>
- <http://www.larioja.org/dato-abierto-rioja/es>
- <http://www.albacete.es/es/webs-municipales/datos-abiertos>
- <http://opendata.unex.es/>
- <http://www.valladolid.es/es/temas/hacemos/open-data-datos-abiertos>
- <http://www.gobiernodecanarias.org/istac/>
- <http://opendataazpeitia.gobernuirekia.net/>
- <http://opendata.gobiernodecanarias.org/opendata>
- <https://sede.ciemat.gob.es/SEDEportal/portal.do?IDM=26&NM=1>
- <http://centrodedescargas.cnig.es/CentroDescargas/>
- http://www.agenciatributaria.es/AEAT.internet/Inicio/La_Agencia_Tributaria/Memorias_y_estadisticas_tributarias/Datos_abiertos_Reutilizacion_de_la_informacion_/Datos_disponibles_catalogo_de_datos_de_la_Agencia_Tributaria/Datos_disponibles_catalogo_de_datos_de_la_Agencia_Tributaria.shtml
- <http://www.aeval.es/datosabiertos/index.html>
- <https://sede.msssi.gob.es/datosabiertos/home.htm>
- https://sede.sepe.gob.es/contenidosSede/generico.do?pagina=datos_abiertos/
- http://www.seg-social.es/Internet_1/Masinformacion/ReutilizacionInformacion/index.htm?ssSourceSiteId=Sede
- <http://datosabiertos.sevilla.org/>
- <http://www.tourspain.es/sites/sedeelectronica/es-es/Paginas/DatosAbiertos.aspx>
- <https://datos.upo.gob.es/>
- <http://opendata.cnmc.es/>
- <http://datosabiertos.laspalmasgc.es/>
- <http://opendataazkoitia.gobernuirekia.eus/>

<https://www.tmb.cat/es/sobre-tmb/herramientas-para-desarrolladores>
<http://opendata.cornella.cat/cornella/ca>
<https://www.aytolucena.es/contenido/datos-abiertos>
<http://www.montilla.es/contenido/datos-abiertos>
<http://www.priegodecordoba.es/contenido/datos-abiertos>
<https://236ws.dpteruel.es/transparencia/dpteruel/open-data/>
<https://seu-e.cat/es/web/castelldefels/dades-obertes>
<https://seu-e.cat/es/web/esparreguera/dades-obertes>
<https://seu-e.cat/web/EspluguesdeLlobregat/dades-obertes>
<https://seu-e.cat/web/igualada/dades-obertes>
<https://seu-e.cat/web/olot/dades-obertes>
<https://sede.guiadeisora.es/publico/openData/catalogo>
<https://seu-e.cat/es/web/amposta/dades-obertes>
<https://sede.teguise.es/publico/opendata/catalogo>
<http://datosabiertos.regiondemurcia.es/>

9.2. CUESTIONARIO A LOS REUTILIZADORES

Cuestionario realizado a los responsables de portales de datos en los meses de marzo y abril de 2017.

- 1.- ¿Sabe de alguna entidad que reutilicen los datos que publican en su portal de datos abiertos? Empresas, ONG, académicos, estudiantes, ciudadanos individuales
- 2.- Indique en qué medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada / nunca..y 5: todo / siempre) [2.a Ciudadanos individuales]
- 2.- Indique en qué medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [2.b Reutilizadores profesionales con ánimo de lucro (infomediarios y desarrolladores individuales con ánimo de lucro)]
- 2.- Indique en qué medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [2.c Reutilizadores profesionales sin ánimo de lucro (ONG, fundaciones, desarrolladores individuales sin ánimo de lucro y otras iniciativas sociales)]
- 2.- Indique en qué medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [2.d Investigadores y personal académico (incluidos estudiantes)]

- 2.- Indique en que medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [2.e La propia organización que publica los datos (otras áreas o departamentos)]
- 2.- Indique en qué medida cada tipo de estas entidades reutilizan sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [2.f Otras entidades públicas distintas de la que publica los datos]
- 3.- ¿A qué sector pertenecen estas entidades? Elija las 5 que más reutilicen [5.- A diario]
- 3.- ¿A qué sector pertenecen estas entidades? Elija las 5 que más reutilicen [4.- Frecuentemente]
- 3.- ¿A qué sector pertenecen estas entidades? Elija las 5 que más reutilicen [3.-Algunas veces]
- 3.- ¿A qué sector pertenecen estas entidades? Elija las 5 que más reutilicen [2.- Excepcionalmente]
- 3.- ¿A qué sector pertenecen estas entidades? Elija las 5 que más reutilicen [1.- Nunca]
- 4.- Indique en qué medida (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), los reutilizadores operan en los siguientes ámbitos [Local (ciudad)]
- 4.- Indique en qué medida (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), los reutilizadores operan en los siguientes ámbitos [Autonómico]
- 4.- Indique en qué medida (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), los reutilizadores operan en los siguientes ámbitos [Nacional]
- 4.- Indique en qué medida (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), los reutilizadores operan en los siguientes ámbitos [Europeo]
- 4.- Indique en qué medida (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), los reutilizadores operan en los siguientes ámbitos [Otros]
- 5.- Indique en qué medida su entidad ha realizado algún tipo de actividad de promoción del uso de sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Concursos de aplicaciones]
- 5.- Indique en qué medida su entidad ha realizado algún tipo de actividad de promoción del uso de sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Actos propios (organizados por su entidad) de presentación de los datos disponibles]
- 5.- Indique en qué medida su entidad ha realizado algún tipo de actividad de promoción del uso de sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Actos externos (organizados por otras entidades) de presentación de los datos disponibles]

- 5.- Indique en qué medida su entidad ha realizado algún tipo de actividad de promoción del uso de sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Reuniones con reutilizadores (de cualquier tipo)]
- 5.- Indique en qué medida su entidad ha realizado algún tipo de actividad de promoción del uso de sus datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Otros (indicar en las notas al final de la encuesta)]
- 6.- Indique en qué medida se produce estos tipos de innovaciones por la reutilización de datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Innovación producto]
- 6.- Indique en qué medida se produce estos tipos de innovaciones por la reutilización de datos (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre), [Innovación proceso]
- 7.- Indique en qué medida, la reutilización de datos ha creado valor en los siguientes ámbitos para las entidades que utilizan los datos o para sus clientes / usuarios (escala 1-5, siendo 1:nada/nunca..y 5: todo/siempre) [Clientes más satisfechos]

9.3. RESPUESTA SOBRE LOS TRES USOS MÁS COMUNES QUE LOS REUTILIZADORES DAN A LOS DATOS DE SU PORTAL

USOS REUTILIZADORES	VALOR
Actividades culturales	1
Algún periódico para publicar información de gastos	2
Aparcamientos	3
Búsqueda de información	4
Consulta información	4
Información en tiempo real: transporte público, gestión de servicios públicos	4
Información de tiempo real	4
Consultas datos turísticos	4.1
Información turísticas	4.1
Consultas de presupuesto público	4.2
Información presupuestaria	4.2
Consultas estructura administrativa	4.3
Información académica	4.4
Información del tráfico en tiempo real: navegadores, rutas	4.5
Ofrecer información de transporte	4.5
Información ranking	4.6

USOS REUTILIZADORES	VALOR
Cartografía	5
Cartografía, planeamiento urbano, geoposicionamiento	5
Utilización de la cartografía para diversos usos	5
Completar bases de datos propias	6
Comunicación ciudadana	7
Informar a la ciudadanía	7
Conocer la agenda de actividades culturales y de deportivas de la ciudad	8
Contratación de servicios técnicos	9
Creación de mapas	10
Dar cumplimiento a la Ley de Transparencia	11
Datos de planificación de la ciudad	12
Datos de planificación de transporte	13
Planificador de rutas	13
Desarrollo de aplicaciones móviles	14
Realización de aplicaciones	14
Estadística de datos	15
Estado de los servicios: playas, parques, transporte público	16
Estudios de mercado	17
Gestión de puntos de interés	18
Horarios de autobuses	19
Labores de periodismo de datos	20
Licitaciones	21
Medioambiental	22
Movilidad	23
Ofertas de empleo	24
Ordenación del Territorio	25
Precios de productos hortofrutícolas	26
Presupuestos	27
Reutilización de la información por la propia organización	28
Transformar su propia organización	29
Valor añadido para aplicaciones o servicios ya existentes	30

9.4.

RESPUESTAS A LOS TRES JUEGOS DE DATOS QUE MÁS SE USAN EN SU PORTAL

JUEGOS DATOS	VALOR
Actividades Culturales y de Ocio	1
Agenda	2
Agenda cultural	2
Agenda diaria	2
Agenda turística	2
Información de agenda cultural	2
Alcaldes de los municipios de la provincia de Alicante. Legislatura 2015	3
Autobuses	4
Bus	4
GTFS de horarios de paso de autobuses	4
Horario de guaguas del Cabildo de La Palma	4
Relación de líneas	4
Routers de transporte público (bus y taxi)	4
Tiempo de llegada de los autobuses	4
TITSA -- Información sobre sistema de transporte.	4
Bici	5
Calendario laboral	6
Calendarios laborales	6
Calidad del aire	7
Cartografía	8
Cartografía escala 1:25.000	8
Información cartográfica	8
Consulta de horarios y frecuencia	9
Contratación de servicios técnicos	10
Contratos adjudicados	11
Convocatorias de personal	12
Ofertas de empleo	12
Relación de puestos de trabajo	12
Demográficos	13
Entidades locales menores y pedanías de la Provincia de Alicante	14
Estado de las playas	15
Playas de la provincia de Cádiz	15
Estado de morosidad	16

JUEGOS DATOS	VALOR
Información presupuesto público	17
Presupuesto	17
Presupuesto. Gastos por programa.	17
Presupuestos	17
Infraestructuras administrativas (centros educativos, oficinas agrarias, oficinas de registro, ...)	18
Listado de asociaciones	19
Mapas	20
Matrículas	21
Modelo digital de elevaciones	22
Movilidad	23
Puntos de recogida selectiva	24
Ranking	25
Recursos educativos	26
Resultados de elecciones sindicales en la propia administración	27
Senderos de la provincia de Cádiz	28
Servicios y equipamientos municipales	29
Tráfico	30

9.5. CARACTERÍSTICAS BÁSICAS DE UN SISTEMA DE GESTIÓN DE DATOS (DMS)

Se incluyen a continuación algunas de las necesidades básicas que cubren los sistemas de gestión de datos (DMS).

Necesidad 1: Publicación de datos. Para poder cubrir esta necesidad el sistema tiene que publicar datos tan pronto como se incorporen los datos y se tengan los permisos adecuados. Permisos, por ejemplo, del propietario de los datos y del administrador de la herramienta. Tiene que permitir así mismo incluir una serie de metadatos¹² cualificando el juego de datos (propietario, frecuencia actualización, última actualización, temática, etiquetas, etc.). Cada juego de datos publicado tiene que rellenar estos metadatos y poder establecer grupos de metadatos que apliquen a varios juegos de datos.

Así mismo tiene que permitir que desde el formato original se puedan generar de forma automática otros formatos técnicos y que puedan ser de utilidad para los reutilizadores. Por ejemplo, si el formato original es un xls, que lo exporte en csv, XML, etc.

Necesidad 2. Búsqueda de datos. Una vez publicados, el número de juegos de datos que pueden llegar a miles, dificulta la localización de forma secuencial. Así debe ser posible la búsqueda por los metadatos que describan el juego de datos u otras características. Así mismo, esta búsqueda debe ser concatenable

y permitir que varios criterios puedan ser aplicados de forma simultánea. Además, y dado que los sistemas de gestión de datos suelen ser públicos en internet, un sistema de gestión de datos también debería proporcionar facilidades para la optimización para motores de búsqueda (SEO).

Necesidad 3: Recopilación de datos. Los datos deben ser actualizados con diversa frecuencia desde la fuente original. Los sistemas de gestión de datos deben incluir facilidades para que esta recuperación se realice de forma automática, idealmente cuando ocurra cualquier cambio. Estos sistemas tienen que ser capaces de crear procesos que hagan de esta tarea un elemento automático y transparente al usuario. En caso contrario, la actualización manual de los datos impondrá una penosa actividad a la organización que demandará recursos adicionales de la misma.

Necesidad 4: Gestión de permisos y flujo de aprobaciones. Algunos de los campos de un juego de datos pueden tener elementos donde la privacidad o la seguridad impidan su publicación. Las herramientas de gestión de datos deben proporcionar un sistema de permisos que dé el control de la publicación a las personas responsables de la organización. La eliminación de estos campos puede ser una solución directa y automatizada, pero también pueden incluirse permisos que eviten que los datos comprometidos sean visibles al público pero que si lo sean para el interior de la organización para aquellas personas autorizadas. Cuanto mayor sea la granularidad de este sistema, mejor.

Necesidad 5: Sincronización de usuarios. Las organizaciones que publican datos son dinámicas y, por tanto, se producen cambios de personas, de roles, etc. Una sincronización de los usuarios permite que estos cambios sean inmediatamente aplicados en los flujos de aprobaciones de la herramienta y así sean transparentes de cara a la publicación de datos. En caso contrario se requerirá de sincronizaciones periódicas o bien actuaciones manuales, que conllevarán mayor uso de recursos.

Necesidad 6: Visualización. Aunque los datos sin procesar pueden ser útiles para muchos de los usuarios, especialmente los de carácter más industrial, los ciudadanos de la calle pueden requerir alguna ayuda para entender su significado, y que se representen en forma de diagramas, mapas, gráficos, etc. Especialmente útil en el caso de que no solo la organización que publica los datos pueda personalizar sino también los propios usuarios mediante un interfaz amigable y que no requiera conocimientos técnicos.

Necesidad 7: Procesado de datos. Aunque idealmente los datos serán recuperados en el formato preciso para ser publicados, la realidad muestra que esta situación es poco frecuente. La posibilidad de filtrar, transformar, eliminar, etc. alguno de los datos originales mediante criterios establecidos permite el aumento de la calidad y de la cantidad de datos publicables.

Necesidad 8. Interfaz de programación de aplicaciones y otros accesos automatizados. Un Interfaz de programación de aplicaciones permitirá un uso industrial y masivo de los datos publicados. Así mismo, reducirá las necesidades de recursos (ancho de banda y capacidad de proceso) al permitir acceder a los datos de forma individual y sin necesidad de

descargar todo el conjunto de datos. Este interfaz debe activarse de forma automática tan pronto como los datos sean publicados. Así mismo, debe estar documentado y, si es posible, proporcionar segmentos de código de ejemplo de cómo acceder a los datos. Finalmente, también podría incluir limitaciones de flujo para aquellos usuarios que ‘abusaran’ del acceso a los datos.

Necesidad 9. Canal de retroalimentación. La publicación de los datos es un proceso continuo, y así una herramienta de gestión de datos debe proporcionar facilidades para que los usuarios puedan informar de incidencias, errores, dar opinión, proporcionar ejemplos de uso, etc. Para que el canal sea efectivo debe incluirse una conexión con la organización que publique los datos e incluir algún tipo de mecanismos de moderación para evitar spam u otras incidencias.

Necesidad 10. Análisis de los accesos. Los sistemas de gestión de datos tienen que incluir elementos que permitan analizar el número y tipo de las descargas realizadas sobre los datos. Y también sobre los usuarios, internos y externos, sobre el número de datos publicados, sobre los comentarios recibidos, etc.

Necesidad 11. Acceso en tiempo real a los datos. Aunque la recopilación de datos que se describió en el punto 3 ya identifica la necesidad de captura periódica de los datos, hay una necesidad creciente de que sea posible la conexión en tiempo real con los sistemas que generan la información. Esta conexión impone necesidades cualitativamente distintas sobre los sistemas de gestión de datos y es progresivamente más demandada por los publicadores. Gradualmente los juegos de datos estáticos versionados se están convirtiendo en chorros de datos también denominados *dataJets* (Abella, 2017).

9.6. CLASIFICACIÓN DE ESTÁNDARES TÉCNICOS SEGÚN MELODA VERSIÓN 4

MELODA establece cuatro niveles de estándares técnicos en cuanto a su fomento de la reutilización.

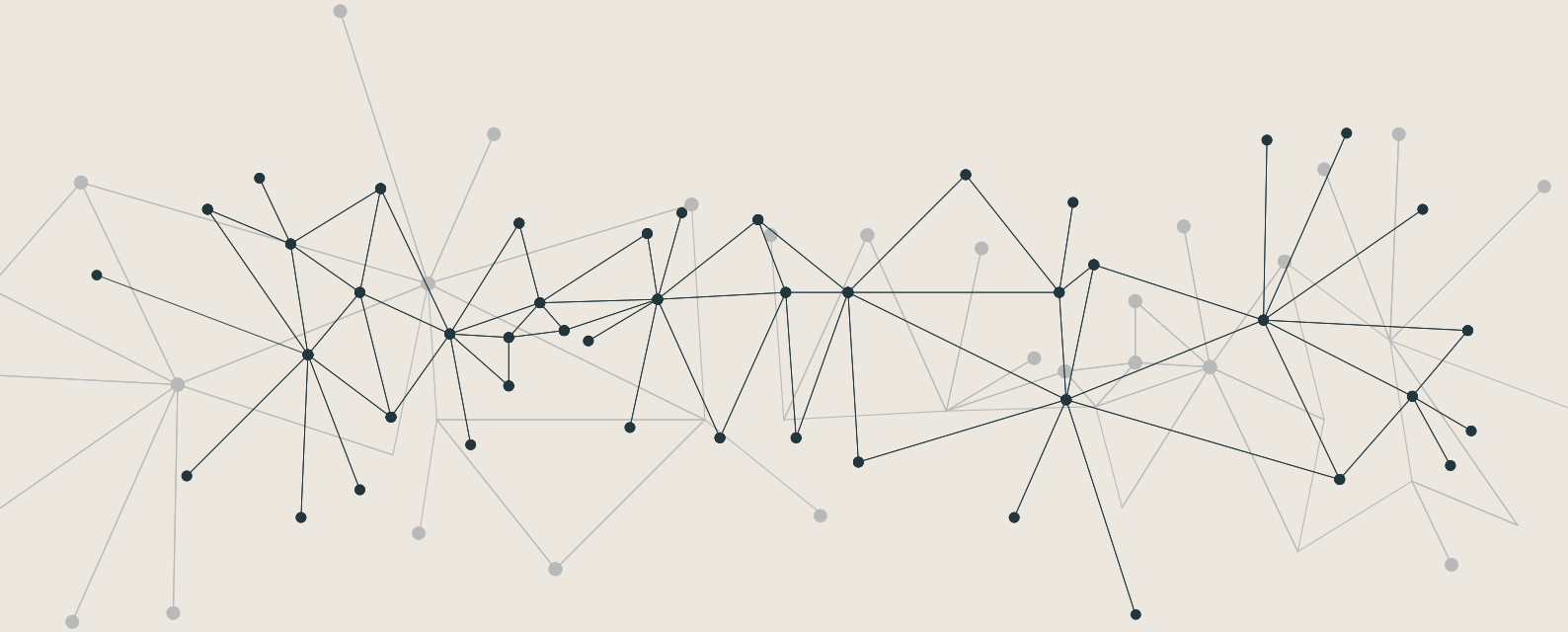
El nivel 1 son aquellos estándares cerrados que requieren herramientas propietarias para su uso. Por ejemplo, un PDF en formato imagen.

En el nivel 2 están aquellos estándares cerrados reutilizables, por ejemplo, XLS y estándares abiertos que no permiten su reutilización como el PDF en formato texto.

En el nivel 3 son estándares abiertos reutilizables (y por tanto ya datos abiertos), por ejemplo, csv, que no incluyen metadatos¹³ en su estructura de almacenamiento.

En el nivel 4 son estándares abiertos que incluyen metadatos, es decir, están aquellos estándares que además de proporcionar los datos incluyen metadatos que proporcionan mayor valor a los datos. Un ejemplo es el formato RDF o XML.

13. Los metadatos a los que se refieren los niveles 3 y 4 de MELODA son los que se incluyen en los formatos técnicos en los que se almacenan los datos.



ic