

---

# *Open data*

---

PID\_00247343

Jesús Alonso-Zárata  
Jordi Casas Roma

---

Tiempo mínimo de dedicación recomendado: 3 horas

---



Universitat  
Oberta  
de Catalunya

---

*Ninguna parte de esta publicación, incluido el diseño general y la cubierta, puede ser copiada, reproducida, almacenada o transmitida de ninguna forma, ni por ningún medio, sea éste eléctrico, químico, mecánico, óptico, grabación, fotocopia, o cualquier otro, sin la previa autorización escrita de los titulares del copyright.*

# Índice

<b>Introducción</b> .....	5
<b>Objetivos</b> .....	6
<b>1. Introducción y definición</b> .....	7
<b>2. Interoperabilidad de los datos</b> .....	8
<b>3. Ventajas y oportunidades de <i>open data</i></b> .....	9
<b>4. Publicar datos en abierto en cinco pasos</b> .....	11
4.1. Identificar los datos que hay que compartir .....	11
4.2. Seleccionar el formato de los datos que hay que compartir .....	12
4.3. Elegir licencia abierta adecuada .....	14
4.4. Ofrecer acceso a los datos .....	14
4.5. Dar visibilidad a los datos abiertos .....	16
<b>5. Decálogo de buenas prácticas compartiendo datos</b> .....	18
<b>6. Datos enlazados: <i>linked data</i></b> .....	20
6.1. Introducción y definiciones .....	20
6.2. El modelo de 5 estrellas de Tim Berners-Lee .....	21
6.2.1. Nivel 1 estrella .....	21
6.2.2. Nivel 2 estrellas .....	22
6.2.3. Nivel 3 estrellas .....	23
6.2.4. Nivel 4 estrellas .....	24
6.2.5. Nivel 5 estrellas .....	26
<b>7. Ejemplos de <i>open data</i></b> .....	28
7.1. Aplicaciones de <i>mashup</i> .....	28
7.2. OpenData BCN: El portal <i>open data</i> del Ayuntamiento de Barcelona	29
7.3. Otros ejemplos .....	30
<b>Resumen</b> .....	32



## Introducción

Este material pretende constituir una introducción al concepto de datos abiertos (del inglés *open data*).

El concepto es sencillo: poner a disponibilidad de uso público datos recogidos para su posterior explotación.

En el caso de la industria 4.0, este concepto de «apertura de datos» puede tener dos claras aplicaciones:

- 1) Generar datos para hacerlos disponibles para terceras partes.
- 2) Usar datos disponibles para ofrecer servicios y optimizar nuestras aplicaciones de valor añadido.

A continuación, veremos una introducción al concepto de datos abiertos, sus posibilidades y algunos ejemplos concretos existentes que nos pueden servir de inspiración.

## Objetivos

Los objetivos de este material son:

- 1) Conocer el concepto de *open data* (datos abiertos) y su aplicabilidad en la visión de la industria 4.0.
- 2) Entender la necesidad de definir estándares de formato de datos para permitir la compatibilidad entre diferentes fuentes de datos.
- 3) Conocer las ventajas e inconvenientes de los datos abiertos.
- 4) Conocer los pasos fundamentales para publicar datos en abierto.
- 5) Familiarizarse con las bases de las buenas prácticas asociadas al uso y generación de datos abiertos.
- 6) Descubrir el potencial de los datos entrelazados y sus aplicaciones en la industria 4.0.
- 7) Estudiar varios casos concretos que aprovechan el uso de los datos abiertos.

## 1. Introducción y definición

La definición de *open data*, o datos abiertos, no es única en la actualidad. Aunque existen distintas aproximaciones con ligeros matices, en esencia no presentan diferencias importantes entre ellas.

La organización Open Knowledge Foundation, en su libro *Open Data Handbook*, define los datos abiertos de la siguiente manera:

«Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requisito de atribución y de compartirse de la misma manera en que aparecen.»

Según la Open Knowledge Foundation, los detalles más importantes en relación con el concepto de datos abiertos se resumen en los siguientes tres puntos:

- 1) **Disponibilidad y acceso:** la información debe estar disponible como un todo y a un coste razonable de reproducción, preferiblemente descargándola de internet. Además, la información tiene que estar disponible en una forma conveniente y modificable.
- 2) **Reutilización y redistribución:** los datos deben ser provistos bajo términos que permitan reutilizarlos y redistribuirlos, e incluso integrarlos con otros conjuntos de datos.
- 3) **Participación universal:** todos deben poder utilizar, reutilizar y redistribuir la información. No debe haber discriminación alguna en términos de esfuerzo, personas o grupos. No se permiten restricciones «no comerciales», que prevendrían el uso comercial de los datos, o restricciones de uso para ciertos propósitos (por ejemplo, solo para educación).

## 2. Interoperabilidad de los datos

El Instituto de Ingenieros Eléctricos y Electrónicos (IEEE) define la **interoperabilidad** como la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar la información intercambiada.

Por lo tanto, la **interoperabilidad de los datos** es la habilidad para interope-  
rar o integrar diferentes fuentes de datos. La interoperabilidad es importante  
porque permite que distintos componentes trabajen juntos. Esta habilidad de  
integrar componentes es esencial para construir sistemas complejos y grandes.

La esencia de los datos abiertos y compartidos es que una parte del material abierto pueda, a partir de ahí, ser mezclado con otro material abierto. Esta interoperabilidad es absolutamente fundamental para entender los principales beneficios prácticos de la apertura: el incremento dramático de la habilidad de combinar distintas fuentes de datos o conjuntos de datos y, así, desarrollar más y mejores productos y servicios.

Proveer una definición clara de apertura garantiza que cuando se trabaje con conjuntos de datos abiertos de fuentes diferentes, se puedan combinar en sistemas más grandes, en los que se encuentra el verdadero valor.

Esta interoperabilidad viene facilitada por las interfaces de programación de aplicaciones (en inglés *application programming interface*), y conocidas por el acrónimo **API**.

Las **API** son el conjunto de funciones y procedimientos que ofrecen cierta fuente estructurada de datos para ser utilizados por un programa externo y poder extraer o almacenar datos.



### 3. Ventajas y oportunidades de *open data*

En la sociedad actual, existen multitud de individuos, organizaciones y, especialmente, administraciones públicas que recopilan y gestionan una gran cantidad y variedad de datos para llevar a cabo sus tareas cotidianas.

En este escenario, las administraciones públicas, como por ejemplo ayuntamientos o gobiernos regionales, son especialmente importantes por la cantidad de datos que manejan, pero también porque una parte importante de esta información es abierta y se pone a disposición de cualquier individuo o institución que desee utilizarla. De este modo, cualquier negocio o industria podría beneficiarse del acceso a estos datos para crear valor añadido a sus productos o servicios.

Existen muchas áreas en las que podemos ver que los datos abiertos han creado valor añadido a la sociedad. A modo de ejemplo, podemos enumerar algunas de las más relevantes:

- Transparencia y control democrático.
- Participación.
- Creación de nuevos productos y servicios.
- Innovación.
- Mejoras en la eficiencia y eficacia de los servicios ofrecidos por el ayuntamiento.
- Medición del impacto de políticas.

Es muy difícil predecir con exactitud cómo, cuándo y dónde se creará el valor a partir de los datos en el futuro. En muchas ocasiones, este valor añadido a la sociedad proviene de los lugares más inverosímiles y de conexiones entre datos que, *a priori*, no parece que tuviesen conexión.

En términos de transparencia, proyectos como Open Spending permiten hacer público de qué manera gobiernos locales y regionales de múltiples países del mundo gastan el dinero de los impuestos de los ciudadanos.

En términos económicos, los datos abiertos son también de gran importancia. Numerosos estudios estiman el valor económico de los datos abiertos en varias decenas de billones de euros al año, solo en la Unión Europea. Cada día más, nuevos productos y compañías están reutilizando datos públicos. Los actores de la industria 4.0 no serán una excepción.

#### Información adicional

Se puede encontrar más información sobre este concepto en el siguiente enlace: <https://openspending.org/>.

Por ejemplo, el traductor de Google usa el enorme volumen de documentos de la Unión Europea que aparecen en todos los idiomas europeos para entrenar sus algoritmos de traducción y, así, mejorar la calidad de su servicio.

Los datos abiertos también pueden contribuir a facilitar las acciones más elementales y cotidianas en la vida de los ciudadanos.

Un ejemplo interesante se puede encontrar en la aplicación FindToilet (disponible para Android y para iOS), que muestra todos los baños públicos daneses que se pueden encontrar a un cierto radio de ubicación actual, usando datos abiertos publicados por el propio Gobierno. Este portal fue creado especialmente para personas con problemas de vejiga, aunque se puede beneficiar del mismo cualquier ciudadano o visitante.

## 4. Publicar datos en abierto en cinco pasos

Cuando una empresa, organización o institución desea publicar datos en abierto, debe plantearse cinco pasos principales que la conducirán a una correcta publicación de los datos en abierto:

- 1) **Identificar** el conjunto o conjuntos de datos que deseamos publicar.
- 2) **Seleccionar** uno o más formatos de datos en los que queremos publicar los datos en abierto.
- 3) **Elegir** una licencia abierta adecuada a los datos que deseamos publicar.
- 4) **Ofrecer** accesibilidad a los datos en un formato que resulte útil para el resto de la comunidad.
- 5) **Dar** visibilidad a los conjuntos de datos que hemos publicado en abierto a través de la comunicación externa.

A continuación, veremos con un poco más de detalle cada uno de los pasos comentados.

### 4.1. Identificar los datos que hay que compartir

El primer paso en el proceso de publicación de datos en abierto es elegir el conjunto o los conjuntos de datos que planeamos abrir.

Este proceso es iterativo, y podemos incluir nuevos conjuntos de datos en el futuro.

En general, no hay requisitos para crear una lista completa de conjuntos de datos que sean candidatos a ser publicados.

Existen **dos puntos principales** que se deben tener en cuenta:

- 1) En primer lugar, debemos asegurarnos de que es factible publicar todos (o una parte) de los datos. Entre otras cosas, es preciso confirmar que no hay datos protegidos o privados cuya publicación pudiera vulnerar alguna ley o protección de propiedad intelectual (*copyright*, patentes, etc.), por ejemplo.
- 2) En segundo lugar, y no menos importante, es necesario asegurarse de que no haya datos personales o privados de personas individuales en el conjunto de los datos

que deseamos publicar. Generalmente, se publican conjuntos de datos que no contienen datos de carácter personal. En caso contrario, deberemos aplicar ciertos procesos de anonimización y preservación de la privacidad que nos garanticen que los datos personales estarán correctamente protegidos en el conjunto de los datos abiertos.

#### 4.2. Seleccionar el formato de los datos que hay que compartir

En segundo lugar, es importante elegir un formato adecuado para la publicación de los datos abiertos. El formato dependerá de varios factores, pero en primer lugar veremos los tipos básicos de estructura o formato de datos que podemos utilizar.

La estructura de datos se define como la forma en que se encuentran organizados un conjunto de datos. Existen diferentes formatos y estructuras con los que podemos representar un mismo conjunto de datos. Clasificaremos los datos, según su nivel de estructuración, en:

- 1) Datos estructurados.
- 2) Datos semiestructurados.
- 3) Datos no estructurados.

1) En los **datos estructurados**, la información viene representada por un conjunto o agrupación de datos atómicos elementales, es decir, datos simples que no están compuestos de otras estructuras. Se conoce de antemano la organización de los datos, la estructura y el tipo de cada dato elemental, su posición y las posibles relaciones entre ellos. Los datos estructurados son de fácil interpretación y manipulación. Los ficheros con una estructura fija en forma de tabla, como los ficheros CSV o las hojas de cálculo, son claros ejemplos de orígenes de datos estructurados.

2) En los **datos semiestructurados**, la información viene representada por un conjunto de datos elementales, pero a diferencia de los datos estructurados, no tienen una estructura fija, aunque poseen algún tipo de estructura implícita o autodefinida. Ejemplos de este tipo de datos son, por ejemplo, los documentos XML o las páginas web. En los dos casos, los documentos siguen ciertas pautas comunes, pero sin llegar a un nivel de estructuración fija.

3) En los **datos no estructurados**, la información no aparece representada por datos elementales, sino por una composición cohesionada de unidades estructurales de nivel superior. La interpretación y manipulación de estos datos resulta mucho más compleja que la de los estructurados o semiestructurados. Ejemplos de orígenes de datos no estructurados son textos, audios, imágenes o vídeos.

Algunos de los tipos de archivos más utilizados para la publicación de datos son:

- Fichero **XLS** o **XSLX**. Es un **formato estructurado** propietario de Microsoft Office para la hoja de cálculo Microsoft Excel; se utiliza en tareas financieras y contables.

- Fichero **de valores separados por comas (CSV)\***. Es un tipo de **documento estructurado** en formato abierto que permite representar datos en forma de tabla, en el que las columnas se separan por comas y las filas, por saltos de línea. Existen variaciones del mismo formato en las cuales las columnas se separan utilizando otros caracteres, por ejemplo, tabulaciones (TSV)\*\*.
- Fichero **XML**. Un fichero XML\* es un tipo de **documento semiestructurado**, compuesto por datos elementales pero de definición previamente no conocida, que incluye etiquetas para describir su propia definición.
- Fichero **JSON\***. Es un **estándar abierto**, basado en texto diseñado para el intercambio de datos legible por humanos, que permite representar estructuras de datos simples y listas asociativas. Es del tipo **semiestructurado**.
- Fichero **RDF\***. Es un **fichero semiestructurado**, basado en sintaxis XML que proporciona los elementos básicos para la descripción de vocabularios.
- Fichero **KML\***. Se trata de un lenguaje de marcado basado en XML para representar datos geográficos en tres dimensiones.
- Fichero **PDF\***. Es un **formato no estructurado** de almacenamiento para documentos digitales multiplataforma que pueden incorporar texto, imágenes vectoriales y mapas de bits.

\* CSV, del inglés *comma-separated values*.  
\*\* TSV, del inglés *tab-separated values*.

\* XML, del inglés *extensible markup language*.

\* JSON, *JavaScript object notation* en inglés.

\* RDF, del inglés *resource description framework*.

\* KML, del inglés *keyhole markup language*.

\* PDF, formato de documento portátil o, en inglés, *portable document format*.

El tipo de fichero que debemos utilizar para generar el conjunto de datos abiertos depende, en gran medida, del tipo de datos que deseamos publicar.

Si deseamos publicar datos en formato de tabla, entonces es aconsejable emplear un tipo de fichero estructurado, como por ejemplo XLS o CSV. Por el contrario, si deseamos publicar datos con una cierta estructura pero que nos permitan flexibilidad, una opción interesante pueden ser los formatos semiestructurados, como por ejemplo XML o JSON. Finalmente, si los datos que deseamos publicar se basan en la localización o puntos de coordenadas GPS, podemos emplear el formato KML.

Por otro lado, siempre es aconsejable utilizar formatos de ficheros abiertos. En caso contrario, solo los usuarios que dispongan de la plataforma propietaria podrán cargar y utilizar el conjunto de datos.

### Ejemplos

El **servicio estadístico Eurostat** ofrece más de cuatro mil conjuntos de datos abiertos, actualizados regularmente, y se obtienen, por lo general, en el formato estructurado y abierto TSV.

El **catálogo de datos del Distrito de Columbia** (*District of Columbia Data Catalog*) ofrece conjuntos de datos abiertos en multitud de formatos, como por ejemplo CSV, XLS o KML, y que además permite acceso a los mismos datos a través de una API abierta.

### 4.3. Elegir licencia abierta adecuada

Una vez que hayamos seleccionado el conjunto o conjuntos de datos y hayamos elegido su formato de publicación, debemos elegir una licencia abierta para su publicación. Es importante seleccionar y aplicar una licencia que establezca de manera clara los usos posibles de los datos publicados.

En este contexto, hay dos atributos de las licencias abiertas que son de especial importancia para seleccionar aquella que mejor se adapte a nuestras necesidades:

- **Atribución (*BY, attribution*)**: indica que el conjunto de datos solo puede ser reutilizado si se reconoce la autoría original en la nueva publicación.
- **Compartir igual (*SA, share-alike*)**: indica que el conjunto de datos solo puede ser reproducido o utilizado como base para la creación de un nuevo conjunto de datos si también se hace bajo una licencia abierta.

Existen multitud de licencias aplicables a datos o conjuntos de datos. Se pueden encontrar diferentes definiciones en las siguientes páginas web:

- <http://opendefinition.org/licenses/>
- <http://opendatacommons.org/guide/>

### 4.4. Ofrecer acceso a los datos

Para cerciorarnos de que los datos abiertos son realmente «abiertos», hay que asegurarse de que enfocamos correctamente dos puntos clave:

- 1) Los datos deben ser abiertos desde un **punto de vista legal**.
- 2) Los datos tienen que ser abiertos desde un **punto de vista técnico**; es decir, debemos facilitar que sean accesibles de forma sencilla y, preferiblemente, que sean datos **legibles por una máquina**. Los datos *legibles en formato máquina*\* son aquellos datos que pueden ser leídos y procesados de forma automática por un computador. Por ejemplo, los documentos de hojas de cálculo son legibles en formato máquina, mientras que una imagen o un documento PDF necesitan un proceso específico para que una máquina pueda interpretar la información contenida en los mismos. Para ilustrar la importancia de la accesibilidad de los datos, supongamos que nuestra organización publica ciertos datos estadísticos en abierto como un documento PDF. Es cierto que este formato favorece la lectura por parte de las personas, pero dificulta tremendamente el hecho de que una máquina pueda leer y entender los datos contenidos en el documento. Por lo tanto, limita en gran medida la reutilización de los datos por parte de otras organizaciones.

\* En inglés, *machine-readable format*.

Existen muchas alternativas para hacer que nuestros datos estén disponibles para otras organizaciones de manera rápida y eficiente. La más natural es la publicación en línea a través de internet. En su forma más básica, las instituciones u organizaciones pueden publicar en abierto sus datos a través de sus propios sitios web. Sin embargo, cuando la conectividad es limitada o el tamaño de los datos es extremadamente grande, la distribución a través de otros formatos puede presentar múltiples ventajas. A continuación, veremos algunas de las alternativas más relevantes y habituales:

- **A través del sitio web de la institución u organización.** Como hemos comentado, esta es la opción más elemental y sencilla de implementar en muchos contextos. Generalmente, los costes de almacenamiento de los datos y del tráfico generado por las descargas de los usuarios son muy bajos, por lo que esta es una opción muy interesante para la publicación de datos de un tamaño razonable.
- **A través del sitio web de terceras partes.** Existen repositorios de datos generalistas y también repositorios especializados en distintos campos. Los sitios web de terceras partes pueden ser muy útiles, dado que generalmente facilitan el acceso a una comunidad de personas interesadas y ponen en común distintos conjuntos de datos similares o complementarios. Además, este tipo de plataformas proporcionan una infraestructura adecuada que puede soportar un volumen de descargas importante, a menudo ofrecen análisis e información de uso y acostumbran a ser gratuitas para las agencias del sector público, por lo que no suelen repercutir el coste de acceso en los usuarios que hacen uso de los datos.

#### Ejemplos

El portal AWS Amazon ofrece un repositorio público multidisciplinar con numerosos conjuntos de datos; el proyecto SNAP de la Universidad de Stanford proporciona acceso a conjuntos de datos de todo tipo de redes (sociales, de comunicación, de transportes, etc.); y el portal Research Data Australia agrupa multitud de recursos abiertos sobre humanidades, ciencias sociales, medicina, tecnología, agricultura y un largo etcétera del Gobierno e instituciones australianas.

- **A través de servidores FTP.** Otra alternativa, aunque en cierto desuso actualmente, es proporcionar acceso a los archivos a través del protocolo de transferencia de archivos (FTP). Esta alternativa puede ser viable si el conjunto de datos que deseamos publicar va dirigido a un colectivo de carácter técnico, como desarrolladores de software o científicos.
- **A través de las redes P2P.** Las redes punto a punto (P2P) trabajan dividiendo el coste de distribución de los archivos entre todas las computadoras que acceden a estos archivos. Es decir, en lugar de que los servidores sean los responsables de enviar la información a todos los clientes que desean acceder a la misma, es la propia comunidad conectada a la red P2P la que distribuye partes de la información entre todos los usuarios. Es una alternativa eficiente para la distribución de volúmenes muy grandes de datos.
- **A través de una API.** Los datos pueden ser publicados a través de una interfaz de programación de aplicaciones (API). Estas interfaces han ganado mucha popularidad recientemente. Algunas de las principales redes sociales, como por ejemplo

#### Ejemplos

El Ayuntamiento de Barcelona publica los datos a través de su OpenData BCN; el Gobierno de Aragón, a través del portal Aragón Open Data; y el Gobierno de España, a través del portal Aporta.

#### Ejemplo

Los datos del Astronomical Data Archives Center del National Astronomical Observatory of Japan son accesibles a través de FTP.

Twitter o Facebook, ofrecen acceso a los datos a través de este tipo de interfaces. Estas permiten a los programadores seleccionar qué tipo o parte de los datos desean descargar, en lugar de proporcionar todos los datos en un archivo de grandes dimensiones. Una ventaja importante es que las API suelen estar conectadas a una base de datos que está siendo actualizada en tiempo real. Esto significa que el suministro de información a través de una API puede proporcionar datos actualizados de forma automática; es decir, si queremos publicar datos en un intervalo de tiempo corto, el uso de API nos puede facilitar tremendamente el trabajo, ya que evita el coste de generar y actualizar grandes ficheros continuamente. Por el contrario, hay que considerar que el coste de desarrollo de una API es muy superior al coste de generar un fichero y distribuirlo a través de las opciones anteriores. Por otro lado, también hay que tener en cuenta que en el caso de utilizar una API, toda la descarga de los datos se llevará a cabo a través de este sistema, de modo que los usuarios que hayan obtenido los datos difícilmente podrán redistribuirlos a terceros.

### Ejemplos

El sitio web Socrata alberga más de cien conjuntos de datos de gobiernos y organizaciones no lucrativas de todo el mundo, a través de una API. El propio World Bank proporciona una gran cantidad de conjuntos de datos abiertos sobre desarrollo en distintos países del mundo a través de su API.

## 4.5. Dar visibilidad a los datos abiertos

Para cerrar de manera satisfactoria el círculo de publicación de datos en abierto, es imprescindible que consigamos conectar a los potenciales usuarios con los datos abiertos. En caso contrario, los datos publicados no tendrán la utilidad que se les supone. Por lo tanto, debemos conseguir que los datos abiertos puedan ser descubiertos por toda la comunidad de usuarios potenciales. Actualmente, podemos encontrar una serie de herramientas o sitios web que están diseñados específicamente para dar visibilidad a los datos abiertos. La propia Open Knowledge Foundation nos ofrece dos herramientas que nos permiten dar visibilidad a nuestros datos.

Por un lado, **CKAN** es una herramienta para la gestión y publicación de colecciones de datos. Esta herramienta ha sido utilizada por distintos gobiernos nacionales y locales, instituciones de investigación y otras organizaciones que recogen una gran cantidad de datos. Los usuarios, ya sean ciudadanos, desarrolladores, periodistas o investigadores, entre otros, pueden buscar datos, registrar conjuntos de datos publicados, crear y administrar grupos de conjuntos de datos, y obtener actualizaciones de bases de datos y los grupos que les interesan.

Por otro lado, podemos obtener libre acceso a muchas de las funciones básicas **CKAN** a través del sitio web **DataHub**. Este sitio web facilita que instituciones y organizaciones puedan publicar el material, y que los usuarios de datos puedan encontrar lo que necesitan.



De manera adicional, encontramos decenas de catálogos especializados en diferentes sectores y territorios. Por ejemplo, muchas comunidades científicas han creado un sistema de catálogo de sus campos, ya que compartir los datos que emplean en sus publicaciones puede ser un requerimiento o puede ayudarlas en la comparación de nuevos métodos y técnicas.

## 5. Decálogo de buenas prácticas compartiendo datos

Dentro del mundo de internet, el World Wide Web Consortium (W3C) es el organismo que se encarga de velar por el desarrollo de estándares abiertos, libres e interoperables que aseguren el crecimiento de la web a largo plazo. Esta organización ha llevado a cabo una guía de publicación con pautas sobre cómo han de publicar datos los gobiernos. Igualmente, existen otras iniciativas generadoras de manuales de buenas prácticas o de concienciación en torno a los datos abiertos, como por ejemplo las aportadas por la Sunlight Foundation o por la Open Knowledge Foundation.

Además, dentro del ámbito estatal, la Comunidad Open Data - Reutilización de Información del Sector Público en España está trabajando en la sensibilización hacia las políticas de *open data* y ha generado algunos documentos de interés. Entre ellos destaca el *Decálogo open data*, que es un resumen de buenas prácticas a la hora de afrontar políticas *open data*, y que transcribimos a continuación.

Todos los puntos del decálogo se basan en la premisa de que debe existir una armonización entre todas las administraciones. Todas las iniciativas *open data* deben compartir los mismos principios y definiciones que se listan en el decálogo. Este punto 0 es básico para la interoperabilidad y el aprovechamiento eficiente de las sinergias llevadas a cabo por todos los actores *open data* - RISP.

- 1) **Publicar datos en formatos abiertos y estándares.** Cualquier iniciativa *open data* debería publicar sus conjuntos de datos en formatos abiertos (no propietarios) y que sean adecuados para permitir su reutilización por parte del colectivo reutilizador destinatario.
- 2) **Usar esquemas y vocabularios consensuados.** Además de los formatos abiertos y estándar, la estructura de los datos debería seguir un convenio o unos esquemas definidos, si existieran. Si se crean vocabularios o esquemas de representación de la información específicos, estos se tendrían que exponer públicamente para que el colectivo reutilizador pueda interpretar correctamente la información.
- 3) **Inventariar el catálogo de datos estructurado.** Cualquier iniciativa *open data* debe tener un punto de consulta en el que se incluya un inventario con información descriptiva y técnica sobre los conjuntos de datos que se exponen. Los metadatos que informan sobre cada conjunto de datos deberían seguir una estructura común y estándar. Asimismo, se deberían compartir las taxonomías de temáticas u otras necesarias –por ejemplo, toponimia– para clasificar los conjuntos de datos dentro de los catálogos.
- 4) **Hacer los datos accesibles desde direcciones web persistentes y amigables.** Tanto las fichas de los conjuntos de datos, como la distribución de la propia infor-

mación (volcado en un archivo, API de consulta, RSS, etc.) deberían ser accesibles desde URL (direcciones web) que persistan en el tiempo, y así evitar que se pierdan las referencias en el futuro. Además, deben seguir una estructura homogénea y bien definida, con información legible para que los reutilizadores conozcan o «intuyan» el contenido referido por estas direcciones web.

5) **Exponer un mínimo conjunto de datos relativos al nivel de competencias del organismo y su estrategia de exposición de datos.** Cada Administración que impulse una iniciativa *open data* debería crear una hoja de ruta donde especifique la estrategia de exposición de los conjuntos de datos y sus prioridades. Inicialmente, debería publicar los conjuntos de mayor interés según las competencias del propio organismo.

6) **Cumplir con un compromiso de servicio, actualización y calidad de los datos, manteniendo un canal eficiente de comunicación con el reutilizador.** La Administración debe mantener un mínimo de calidad y servicio en su iniciativa *open data*, manteniendo lo expuesto en la estrategia de publicación y comprometiéndose con su colectivo reutilizador. Debe establecer un canal eficiente de comunicación que permita la interacción bidireccional organismo público - reutilizadores.

7) **Monitorar y evaluar el uso y servicio mediante métricas.** La Administración debe crear métricas y evaluar sus indicadores de uso y servicio de la iniciativa *open data*. De esta manera, puede monitorizar su funcionamiento y uso, y así analizar si se está cumpliendo el compromiso con la comunidad de reutilizadores y cuáles son las potenciales carencias del sistema o de la estrategia.

8) **Ofrecer los datos bajo condiciones de uso no restrictivas y comunes.** Las condiciones de uso deberían ser lo menos restrictivas que se pueda y permitir la reutilización libre, incluso para fines comerciales. Se recomienda la creación y el uso de licencias tipo, autodocumentadas y que sean comunes entre distintas administraciones.

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

10) **Recopilar aplicaciones, herramientas y manuales para motivar y facilitar la reutilización.** Cualquier iniciativa *open data* debería recopilar ejemplos de uso y herramientas que faciliten y motiven la reutilización de los datos que se publican.

## 6. Datos enlazados: *linked data*

### 6.1. Introducción y definiciones

Estrechamente vinculado al concepto datos abiertos, se encuentra el concepto de *datos enlazados*\*.

\* En inglés, *linked data*.

El término **datos enlazados** hace referencia al método con el que se pueden mostrar, intercambiar y conectar datos abiertos a través de un URI en la web.

Pero, ¿qué significa URI?

**URI**\* es un identificador de recursos uniforme. Se trata de una cadena de caracteres que identifica los recursos de una red de forma unívoca.

\* Del inglés *uniform resource identifier*.

Aunque la definición completa y detallada de URI queda fuera del alcance de este texto, y requiere ciertos conocimientos técnicos específicos que no son relevantes para este material, la estructura básica de un URI es la siguiente:

esquema: //máquina/directorio/archivo#fragmento

Un posible ejemplo sería este:

```
http://ejemplo.com/ruta/data?key=value\#fragmento1
```

Los datos enlazados presentan un método de publicación de datos estructurados para que puedan ser interconectados y, así, usados para generar valor.

El concepto se basa en el uso de tecnologías web estándar, tales como HTTP, RDF y los URI. Sin embargo, en lugar de utilizarlos para servir páginas web que puedan leer personas, estas se extienden para compartir información de una manera que puede ser leída automáticamente por ordenadores. Esto permite conectar y consultar datos de diferentes fuentes de datos. Las posibilidades para la industria 4.0 son realmente ilimitadas.

El marco de descripción de recursos o RDF\* es un método general para el modelado de información. Permite describir metadatos en sitios web, así como proveer **interoperatividad** entre las aplicaciones que intercambian información en lenguaje máquina por la web. Permite detallar información como las fechas de actualizaciones de páginas, palabras clave, derechos de autor, etc.

\* Del inglés *resource description framework*.

## 6.2. El modelo de 5 estrellas de Tim Berners-Lee

Tim Berners-Lee, científico británico en el área de la computación, es conocido por ser el padre de la Web. Estableció la primera comunicación entre un cliente y un servidor usando el protocolo HTTP en noviembre de 1989. En octubre de 1994, fundó el Consorcio de la World Wide Web (W3C) para supervisar y estandarizar el desarrollo de las tecnologías sobre las que se fundamenta la web, y que permiten el funcionamiento de internet.

Tim Berners-Lee sugirió un esquema de desarrollo de 5 estrellas para datos abiertos. Según esta clasificación, los datos abiertos pueden convertirse en datos enlazados si se interrelacionan.

Los datos enlazados son la base técnica para crear la denominada web semántica, un estándar en el que cada dato contiene información asociada que lo relaciona automáticamente con otros datos.

Podemos considerar que los datos abiertos, como los hemos presentado hasta ahora, llegan al nivel de 3 estrellas del esquema de Tim Berners-Lee. A partir del nivel 4 y 5, es necesario añadir la interconexión entre distintos conjuntos para dotarlos de significado.

A continuación, mostramos cada uno de los niveles del esquema de 5 estrellas y comentaremos los principales beneficios involucrados en cada caso. Además, a lo largo de las subsecciones, presentamos un ejemplo práctico que ayude a comprender los conceptos que se plantean.

### 6.2.1. Nivel 1 estrella

El nivel 1 estrella significa que los datos se publican en una web, con cualquier formato de datos, y bajo una licencia abierta. Publicar datos con 1 estrella implica que:

- Podemos verlos.
- Podemos imprimirlos.
- Podemos guardarlos localmente (en un disco duro o en una memoria USB).
- Podemos ingresar los datos en cualquier otro sistema.
- Podemos cambiar los datos como queramos.
- Podemos compartir los datos con quien queramos.

Como ejemplo, supongamos que el Ayuntamiento de Barcelona desea publicar en abierto datos sobre la previsión meteorológica de los próximos días en la ciudad. En un primer momento, el Ayuntamiento decide publicar en su sitio web un documento PDF con la previsión meteorológica de los próximos días bajo una licencia abierta, por ejemplo, la Creative Commons - Attribution 4.0 (CCBY 4.0).

El contenido del PDF puede ser, por ejemplo, una tabla como la que se muestra a continuación.

Figura 1

Previsión meteorológica de Barcelona

Día	Temp. Mín.	Temp. Máx.	Lluvia	Cielo
1/01/2016	7	16	No	Cubierto
2/01/2016	8	18	No	Soleado
3/01/2016	3	11	Si	Cubierto

De este modo, el Ayuntamiento ha conseguido la primera estrella del modelo de Tim Berners-Lee. Es decir, los datos son accesibles a través de internet bajo una licencia abierta.

### 6.2.2. Nivel 2 estrellas

El nivel 2 estrellas implica publicar los datos como datos estructurados (por ejemplo, formato de Microsoft Excel en lugar de una imagen de una tabla escaneada).

Como usuarios de los datos, las ventajas de utilizar datos publicados en el nivel 2 estrellas son las mismas que en el nivel anterior, más las siguientes:

- Podemos procesarlos directamente con software propietario para agregarlos, hacer cálculos, visualizarlos, etc.
- Podemos exportarlos a otro formato (estructurado).

Siguiendo con el ejemplo del Ayuntamiento de Barcelona, el principal problema que tienen en este momento es la dificultad de las terceras partes para procesar esta información. El Ayuntamiento ha publicado la información utilizando un formato no estructurado de datos en un documento PDF, por lo que no será posible su posterior tratamiento. Si el usuario que accede a la información solo desea consultarla, no tiene que haber ningún problema. Sin embargo, en el caso de que desee llevar a cabo algún tipo de transformación (por ejemplo, incluirla en otro documento o insertarla en una base de datos) u operación (por ejemplo, analizar o agregar la información), deberá copiar e introducir manualmente los datos en la aplicación en la que desee utilizarlos.

Si el conjunto de datos es grande, este proceso puede ser tedioso y consumir mucho esfuerzo y tiempo. Para conseguir que los datos alcancen el nivel de la segunda estre-

lla, sería conveniente publicar los datos abiertos en un formato estructurado. En este caso, se podría elegir, por ejemplo, el formato propietario de Microsoft Excel, conocido como XLS, en lugar de un documento PDF. De este modo, los datos tendrían este aspecto:

Figura 2. Ejemplo *open data* con 2 estrellas

	A	B	C	D	E	F
1	Día	Temp. Mín.	Temp. Máx.	Lluvia	Cielo	
2	1/1/16	7	16	No	Cubierto	
3	2/1/16	8	18	No	Soleado	
4	3/1/16	3	11	Si	Cubierto	
5						

De este modo, los datos publicados podrían ser cargados, utilizando un programa propietario, y analizados o procesados de manera automática. Por ejemplo, mediante Microsoft Excel, los usuarios o terceras partes podrían calcular la temperatura mínima media en los próximos días o contar los días con probabilidad de lluvia.

### 6.2.3. Nivel 3 estrellas

Nivel 3 estrellas implica utilizar formatos no propietarios, como por ejemplo CSV en lugar de un formato propietario específico, como podría ser el formato de Microsoft Excel, por poner un ejemplo.

Las ventajas de utilizar datos publicados en el nivel 3 estrellas son las ventajas del nivel 2 estrellas más la posibilidad de manipular los datos de cualquier forma que queramos, sin limitación de las características o uso de un tipo de software en particular. En este caso, es posible que la persona o institución que deseen publicar datos en abierto de nivel 3 estrellas necesiten convertidores o *plug-ins* para exportar los datos desde el formato propietario.

Siguiendo con el ejemplo anterior, el Ayuntamiento de Barcelona podría publicar los datos en formato de fichero separado por comas (CSV). De este modo, los datos cumplirían con la tercera estrella del modelo. Los datos tendrían este formato:

Figura 3. Ejemplo *open data* con 3 estrellas

1	Día,Temp. Mín.,Temp. Máx.,Lluvia,Cielo
2	1/1/16,7,16,No,Cubierto
3	2/1/16,8,18,No,Soleado
4	3/1/16,3,11,Si,Cubierto

De este modo, los datos pueden ser descargados por cualquier ciudadano o institución, cargados en un programa no propietario y analizados o procesados.

#### 6.2.4. Nivel 4 estrellas

El nivel 4 estrellas implica usar URI para apuntar a los datos. A partir de este nivel, hablaremos de datos abiertos y enlazados. El objetivo principal de este nivel es utilizar URI para identificar los conceptos que aparecen en los datos abiertos.

Como usuarios de datos abiertos, podemos hacer todo lo que hacemos en el nivel 3 estrellas con los datos y, además:

- Podemos enlazarlos desde cualquier otro sitio (web o local).
- Podemos marcarlos como favoritos.
- Podemos reutilizar partes de los datos.
- Podríamos reutilizar herramientas y librerías disponibles, incluso si estas solo entienden parte de los patrones que utilizó quien los publicó.
- Podemos combinar nuestros datos con otros conjuntos de datos. Los URI son un esquema global, por lo que si dos cosas tienen el mismo URI, entonces representan el mismo concepto.

Por el contrario, es relevante destacar que entender la estructura de un documento de datos RDF puede requerir más esfuerzo que el necesario para entender datos estructurados tipo XLS o CSV.

Como persona o institución que desee publicar datos en abierto, debemos considerar los siguientes aspectos:

- Otros conjuntos de datos podrán enlazarse a nuestros datos.
- Deberemos asignar URI a los datos y pensar en cómo representarlos.

Seguimos con el ejemplo del Ayuntamiento de Barcelona. Supongamos ahora que se decide publicar los datos en formato RDF para cumplir con el modelo de 4 estrellas.

No incluiremos el código completo del ejemplo en este documento, ya que es largo, complejo y se escapa de los objetivos de este material. Aun así, veremos algunas partes que nos pueden ayudar a entender el contexto semántico que se añade en este nivel de 4 estrellas.

En este ejemplo, utilizaremos el vocabulario de Meteo para dar contexto a ciertos conceptos que se usan en este conjunto de datos del ejemplo. La declaración del vocabulario se hace en la línea 3 del código mostrado en el código de la figura 4.



Figura 4. Ejemplo *open data* con 4 estrellas

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/Markup/DTD/
  xhtml-rdfa-1.dtd">
3 <html xmlns:meteo="http://purl.org/ns/meteo#">
4 <head>
5   <title>Previsión meteorológica de Barcelona</title>
6 </head>
7 <body>
8   <div id="data" about="#Barcelona" typeof="meteo:Place">
9     <table border="1px">
10      <tr>
11        <th>Dia</th>
12        <th>Temp. Min.</th>
13        <th>Temp. Max.</th>
14        <th>Lluvia</th>
15        <th>Cielo</th>
16      </tr>
17      <tr rel="meteo:forecast" resource="#prevision01012016">
18        <td>
19          <div about="#prevision01012016">
20            <span property="meteo:predicted" datatype="xsd:dateTime">01/01/2016</span>
21          </div>
22        </td>
23        <td rel="meteo:temperature">
24          <div about="#temp01012016">
25            <span property="meteo:celsius" datatype="xsd:decimal">7</span>
26          </div>
27        </td>
28        ...
29      </tr>
30      ...
31    </table>
32  </div>
33 </body>

```

Esto nos permite contextualizar ciertos conceptos, como por ejemplo el lugar de donde se está dando la previsión (podemos ver la declaración `typeof="meteo:Place"` y `about="#Barcelona"` en la línea 8, que indica la localización de la previsión) o el tipo de datos que mostramos en cada parte (la línea 25 indica que utilizaremos los grados Celsius para la temperatura `property="meteo:celsius"`).

Además, se asignará un URI distinto a la previsión de cada día de forma independiente. Así, datos externos podrán hacer referencia a la previsión que se ha llevado a cabo de un día concreto. Por ejemplo, en la línea 17 se declara la realización de una previsión (`rel="meteo:forecast"`) y se le asigna un identificador (`resource="#prevision01012016"`), que indica que es la previsión para el día 1 de enero del 2016 (01-01-2016). Si un conjunto externo quiere referenciar esta previsión, deberá utilizar un URI de la forma siguiente:

<http://www.ejemplo.com/datosenlazados/4star/\#prevision01012016>

Donde se supone que los datos se publican en el dominio «[www.ejemplo.com](http://www.ejemplo.com)» y en la ruta de acceso «[datosenlazados/4star/](http://www.ejemplo.com/datosenlazados/4star/)».

Aunque el fichero RDF presente una estructura jerárquica, hay que entenderlo como un conjunto de declaraciones o hechos simples. Por ejemplo, en el código 2 se declara que «[#Barcelona](#)» es «un lugar» y que «[#Barcelona](#)» tiene un conjunto de previsiones, entre ellas «[#prevision01012016](#)».

### 6.2.5. Nivel 5 estrellas

El nivel 5 estrellas implica enlazar los datos publicados con otros datos también publicados en internet, para proveer contexto. Como usuarios de datos abiertos, podemos hacer todo lo que hacemos en el nivel 4 estrellas y además:

- Descubrir más datos (relacionados) mientras consumimos los datos.
- Aprender directamente acerca del esquema de datos.

También deberemos considerar los problemas generados por los enlaces rotos de datos, como los errores de páginas web no encontradas. Como persona o institución que desee publicar datos en abierto, debemos considerar los siguientes aspectos:

- Podemos hacer que nuestros datos sean descubiertos.
- Podemos incrementar el valor de nuestros datos.

En este caso, también tendremos que lidiar con problemas de enlaces rotos o incorrectos.

Seguimos con el ejemplo del Ayuntamiento de Barcelona. El código mostrado en la figura 5 presenta unos datos similares al nivel anterior, en el que se han añadido enlaces a sitios externos que dan contexto a ciertos conceptos que en el nivel anterior no estaban del todo definidos.

Figura 5. Ejemplo *open data* con 5 estrellas

```

1 <?xml version="1.0" encoding="utf-8"?>
2 <!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN" "http://www.w3.org/MarkUp/DTD/
  xhtml-rdfa-1.dtd">
3 <html xmlns:meteo="http://purl.org/ns/meteo#">
4 <head>
5   <title>Previsión meteorológica de Barcelona</title>
6 </head>
7 <body>
8   <div id="data" about="#Barcelona" typeof="meteo:Place">
9     <span rel="owl:sameAs" resource="http://dbpedia.org/resource/Barcelona"></span>
10    <table border="1px">
11      <tr>
12        <th>Dia</th>
13        <th>
14          <a rel="rdfs:seeAlso" href="https://es.wikipedia.org/wiki/Temperatura" resource
15            ="http://dbpedia.org/resource/Temperature">Temp.</a> Min.
16          (<span rel="owl:sameAs" resource="http://dbpedia.org/resource/Celsius">°C</span
17            >)
18        </th>
19        <th>Temp. Max.</th>
20        <th>Lluvia</th>
21        <th>Cielo</th>
22      </tr>
23      <tr rel="meteo:forecast" resource="#prevision01012016">
24        <td>
25          <div about="#prevision01012016">
26            <span property="meteo:predicted" datatype="xsd:dateTime">01/01/2016</span>
27          </div>
28        </td>
29        <td rel="meteo:temperature">
30          <div about="#temp01012016">
31            <span property="meteo:celsius" datatype="xsd:decimal">7</span>
32          </div>
33        </td>
34        ...
35      </tr>
36      ...
37    </table>
38  </div>
39 </body>

```

Veamos un par de ejemplos que nos pueden ayudar a comprender el concepto de contextualización al que nos estamos refiriendo.

- 1) En el nivel anterior de 4 estrellas, se ha definido que «Barcelona» es «un lugar». Sin embargo, en este nivel de 5 estrellas se ha añadido, en la línea 9 del código, una referencia a una página web que trata sobre Barcelona en la DBPedia <http://dbpedia.org>. DBpedia es un proyecto para la extracción de datos de Wikipedia para proponer una versión web semántica. Esta página contiene información estructurada sobre Barcelona, de manera que una máquina es capaz de procesar esta información y enlazar a otros sitios que ayudarán a definir y contextualizar la información de este sitio.
- 2) La línea 14 enlaza el concepto «temperatura» con las páginas de Wikipedia y DBPedia para proporcionar contextualización del concepto para lectura humana y máquina, respectivamente. También se ha añadido información sobre el formato de la temperatura, y se indica que este dato se ha facilitado en grados Celsius (°C), de modo que enlaza con su contexto de DBPedia.

## 7. Ejemplos de *open data*

### 7.1. Aplicaciones de *mashup*

Un claro ejemplo de integración y reutilización de datos lo encontramos en las denominadas *mashup* o aplicaciones web híbridas.

Estas aplicaciones se crean a partir de la reutilización de contenidos y/o funcionalidad de sitios web de terceros que permiten el acceso a datos abiertos.

A partir de distintas fuentes de datos, estas aplicaciones permiten crear nuevos servicios simples, y visualizar en una única interfaz gráfica la combinación de los datos extraídos.

Por ejemplo, se pueden combinar las direcciones y fotografías de una biblioteca con un mapa de Google Maps para crear un *mashup* de mapa.

El término *mashup* implica integración fácil y rápida, a menudo usando varias API abiertas y fuentes de datos para producir resultados enriquecidos, que no fueron necesariamente el motivo original de producir la fuente primaria de datos.

La arquitectura de los *mashups* está compuesta de tres partes principales:

- 1) El proveedor de contenidos o fuente de los datos proporciona los datos abiertos que servirán como información base para crear el servicio que se desea implementar. Generalmente, se accede a los datos a través de una API.
- 2) El sitio *mashup* es la nueva aplicación web que provee un nuevo servicio utilizando diferente información mezclada a partir de distintas fuentes abiertas.
- 3) El navegador web cliente es la interfaz de usuario del *mashup*.

A modo de ejemplo concreto, el Departamento de Policía de Chicago tiene un *mashup*, llamado *Chicago crime*, que integra la base de datos del departamento de crímenes reportados con Google Maps, con el objetivo de ayudar a detener crímenes en ciertas áreas y avisar a los ciudadanos de áreas potencialmente peligrosas.

## 7.2. OpenData BCN: El portal *open data* del Ayuntamiento de Barcelona

El Ayuntamiento de Barcelona dispone del portal OpenData BCN, con datos abiertos relacionados con la ciudad. En la web <http://opendata.bcn.cat/opendata/es/what-is-open-data>, el propio Ayuntamiento define los datos abiertos de la manera siguiente:

«La apertura de datos públicos –también conocido como *open data*– consiste en poner la información que posee el sector público al alcance de todo el mundo en formatos digitales, estandarizados y abiertos, siguiendo una estructura clara que permita su comprensión. Al mismo tiempo, se facilita el acceso a esta información con el fin de fomentar su reutilización.»

El Ayuntamiento publica un catálogo de más de trescientos conjuntos de datos abiertos, clasificados en las siguientes categorías:

- **Administración.** Incluye conjuntos de datos relacionados con temas de legislación, justicia y sector público.
- **Ciudad y servicios.** Dentro de esta categoría, se incluyen todos los conjuntos de datos abiertos que hacen referencia a cultura, ocio, deporte, medio ambiente, seguridad, transporte y turismo de la ciudad.
- **Economía y empresa.** Integra los datos relacionados con ciencia, tecnología, comercio y empleo de la ciudad.
- **Población.** Agrupa los conjuntos de datos relacionados con demografía, educación, sociedad y bienestar.
- **Territorio.** Finalmente, esta categoría contiene datos sobre temas de urbanismo, infraestructuras y vivienda.

Los datos se publican en distintos formatos, la mayoría de los cuales han sido comentados en este material. Los más utilizados corresponden a datos estructurados en formato de tablas, como por ejemplo CSV, XLS, XLSX.

No obstante, también encontramos múltiples documentos en formatos semiestructurados, como JSON o XML; formatos para uso específico de geolocalización, como GEO y KML; y también formatos no estructurados, como ficheros de texto plano (TXT) o documentos PDF.

Todos los conjuntos de datos que se ofrecen en el servicio OpenData BCN (<http://opendata.bcn.cat/>) indican qué licencia y condiciones de uso tienen. La mayoría de los conjuntos de datos se publican bajo los términos de la licencia Creative Commons - Reconocimiento (CC-BY 3.0). Por lo tanto, es posible:

- Copiar, distribuir y divulgar públicamente los datos.

- Usar los datos como base para obras derivadas, como resultado de su análisis o estudio.
- Usar los datos con fines comerciales o no comerciales, siempre que este uso no constituya una actividad administrativa pública.
- Modificar, transformar y adaptar los datos.

Sin embargo, en el caso de tipos de datos en los que hay participación de terceros, la reutilización se vehicula a través de la licencia Creative Commons - Reconocimiento - SinObraDerivada (CC BY-ND 3.0), que permite lo siguiente:

- Que se puedan copiar, distribuir y divulgar públicamente.
- Que puedan ser utilizados con fines comerciales o no comerciales, siempre que este uso no constituya una actividad administrativa pública.
- Que se puedan modificar, transformar y adaptar.
- Que se deba mencionar la autoría de los datos.

Por otro lado, de acuerdo con el artículo 8 de la Ley 37/2007, la reutilización de la información contenida en los conjuntos de datos está sometida a las condiciones generales siguientes:

- Que el contenido de la información no sea alterado.
- Que no se desnaturalice el sentido de la información.
- Que se mencione la fuente.
- Que se mencione la fecha de la última actualización.

Además, el sitio web dispone también de una herramienta que permite la visualización de algunos de los conjuntos de datos sobre un mapa de la ciudad.

El único requisito para su uso es que se debe mencionar la autoría del Ayuntamiento de Barcelona. Se pueden consultar las condiciones de uso completas del portal OpenData BCN en la dirección web <http://opendata.bcn.cat/opendata/es/data-using>.

### 7.3. Otros ejemplos

Existen multitud de servicios que se han creado a partir de iniciativas parecidas de publicación de datos por parte de administraciones públicas. A continuación veremos, a modo de ejemplo, algunos servicios relacionados con distintas temáticas:

- **El transporte y la movilidad de las personas.** Por ejemplo, InfoCarretera, un servicio gratuito que informa, en el móvil, sobre el estado del tráfico en las carreteras y ciudades del País Vasco.
- **Servicios relacionados con la educación.** Un ejemplo es Schoolscope, un sencillo servicio que informa sobre la calidad de la educación en cada una de las escuelas inglesas, y permite contrastar la información.
- **La salud.** Una interesante iniciativa americana es Apps for Healthy Kids, un concurso de servicios pensados para promover una vida saludable entre los niños.
- **La seguridad ciudadana.** Una aplicación americana para móvil muy conocida es AreYouSafe?, que muestra el grado de delincuencia en función del lugar donde uno se encuentra.
- **La transparencia en el sector público.** El servicio «¿Dónde van mis impuestos?» explica, de manera muy visual, los gastos anuales de la Administración central del Estado y la Seguridad Social, tal y como aparecen recogidos en los presupuestos generales.

## Resumen

En este material, hemos introducido el concepto de *open data* (datos abiertos). El hecho de que el mundo vaya a estar hiperconectado permite disponer de una cantidad muy grande de información que, en algunos casos, puede ser compartida de manera abierta, para que otros puedan explotarla.

El concepto de datos abiertos lo podemos ver, por lo tanto, desde dos perspectivas:

- 1) El que genera datos.
- 2) El que usa datos.

Las dos opciones pueden resultar interesantes para actores de la industria 4.0.

En cualquier caso, hemos visto la necesidad de definir estándares de formato de datos que permitan asegurar la compatibilidad entre diferentes fuentes de datos; de este modo, podremos tratarlos y combinarlos para generar valor con los datos.

Hemos visto en el material que la filosofía de datos abiertos ofrece muchas ventajas, pero también inconvenientes que hay que saber equilibrar para decidir, en cada caso, si es una alternativa que hay que considerar para nuestros intereses.

Encontramos cinco pasos que hay que tener en cuenta para compartir datos en modo abierto:

- 1) **Identificar** los datos que queremos compartir.
- 2) **Seleccionar** el formato en el que vamos a compartir datos.
- 3) **Elegir** la licencia abierta que más nos interesa para compartir los datos.
- 4) **Ofrecer** acceso a los datos.
- 5) **Dar visibilidad** a los datos que hemos compartido.

También hemos presentado un decálogo de buenas prácticas en lo que respecta a compartir y usar datos abiertos.

El uso de datos abiertos entrelazados, es decir, en combinación de diferentes fuentes, abre un abanico único de posibilidades para crear valor. Posteriormente, hemos descrito brevemente el modelo de estrella de Tim Berners-Lee, que permite clasificar la interconexión entre datos abiertos.

Finalmente, hemos presentado varios casos concretos que aprovechan el uso de los datos abiertos y que nos pueden servir de inspiración y espejo en el que mirarnos a la hora de decidir optar por compartir datos en modo abierto, o bien usar datos abiertos para poder generar valor.