



Datos abiertos y formatos abiertos

Universidad de Granada: Oficina de Software Libre
Autor: Juan Manuel López Castro

Índice

- 1- Introducción
 - Nivel de madurez
 - Estrategias
 - Impulsores, barreras y beneficios
- 2- Leyes aplicadas al open data
 - Licencias
 - Legislación Europea
 - Normas Técnicas de Interoperabilidad (NTI)
 - NTI de reutilización.
- 3- Protocolos y software de apoyo
 - Open Data Protocol
 - SPARQL
 - DCAT-AP
 - CKAN
 - DKAN
- 4- Formatos abiertos
 - DOCX
 - Formatos en investigaciones
 - Ciencia abierta

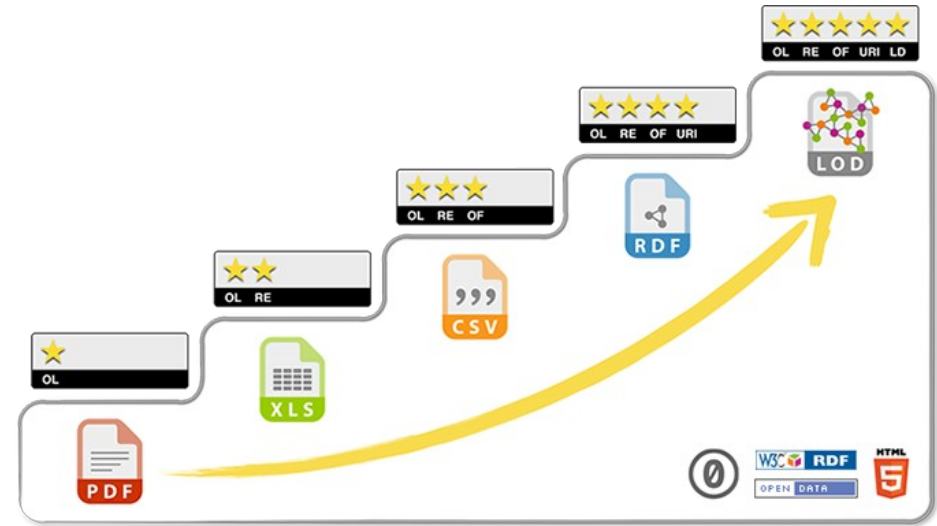
1. Introducción

- El fenómeno de open data es el proceso que pone al alcance de la sociedad los datos generados por las entidades públicas.
- Datos de tráfico, de investigaciones, estadísticos, meteorológicos, etc.
- Deben de cumplir una serie de condiciones:
 - Completos.
 - Puros.
 - Disponibilidad reciente (actualizados).
 - Accesibles para toda la población.
 - Procesables por computadores (digitales y estructurados).
 - No discriminatorios (sin registro o autenticación).
 - Sin propiedad.
 - Licencia gratuita.



1.1 Nivel de madurez

- A la hora de publicar los datos existen diferentes niveles de madurez dependiendo de su formato:
 - Nivel 1: no tienen ningún nivel de estructura (pdf, imágenes, videos, etc)
 - Nivel 2: estructura apoyada por un software.
 - Nivel 3: no se apoya en ningún software para su estructura.
 - Nivel 4: estándar RDF identificado por URIs.
 - Nivel 5: estructurados con URIs y enlazados entre sí (dota a los datos de un contexto).



1.2 Estrategias

Dentro del proceso de open data se pueden encontrar tres estrategias o motivaciones predominantes:

- España, Australia, Reino Unido, Estados Unidos y Dinamarca.
- 1- Aumentar el derecho democrático y la participación política → democracia moderna, estar informados en que se gasta el dinero
- 2- Fomentar la investigación e innovación → nuevas oportunidades al explotar los datos
- 3- Fortalecer la aplicación de la ley → implicando a los ciudadanos



1.3 Impulsores, barreras y beneficios

- Impulsores

- Presión ciudadana.
- Iniciativas llevadas por empresas.
- Nuevas tecnologías.
- Investigaciones.

- Barreras

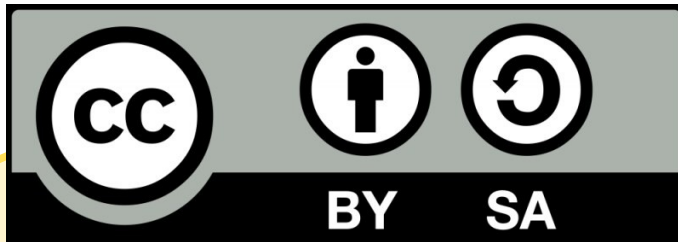
- Culturales (necesita de un gran compromiso político).
- Legales y organizativas.
- Prácticas y técnicas (cuestiones lingüísticas e interoperabilidad).

- Beneficios

- Realización de investigaciones.
- Transparencia.
- Oportunidades a la creación de empresas.
- Valor económico.

2. Leyes aplicadas al Open Data: licencias

- Tipos de licencias
 - Licencias de uso comercial.
 - Licencias de uso no comercial.
 - Licencias abiertas.
 - Licencias previa solicitud.
 - Licencias con condiciones.
- Creative Commons
 - Creative Commons Zero (CCO) → sin ningún tipo de restricción.
 - Creative Commons Reconocimiento (CC BY) → mención sobre la fuente original.
 - Creative Commons Reconocimiento-NoDerivadas (CC BY-ND) → similar a CC BY, añadiendo la no modificación.
 - Creative Commons Reconocimiento-NoComercial (CC BY-NC) → alterar los datos cuando se mencione la fuente y uso no comercial.



2. Leyes aplicadas al Open Data: licencias

- Tipos de licencias
 - Licencias de uso comercial.
 - Licencias de uso no comercial.
 - Licencias abiertas.
 - Licencias previa solicitud.
 - Licencias con condiciones.
- Open Data Commons
 - Open Data Commons Public Domain Dedication and License (PDDL) → permite difundir, reutilizar y adaptar los datos
 - Open Data Commons Attribution License → reconocimiento de la fuente
 - Open Data Commons Open Database License (OdbL) → implica la autoría y el uso de la misma licencia en obras derivadas.

2.1 Leyes aplicadas al Open Data: legislación Europea

- Reforma sobre datos abiertos y reutilización de información DIRECTIVA (UE) 2019/1024 DEL PARLAMENTO EUROPEO.
 - Datos de investigación incluidos en la regulación europea (investigaciones con fondos públicos).
 - Uso obligatorio de metadatos.
 - Datos bien estructurados para una buena reutilización
 - Datos dinámicos de forma accesible y actualizada (APIs).
 - Datos geoespaciales, medio ambientales, movilidad, etc, deben de estar suministrados de forma gratuita.
 - Criterio de la mera recuperación de los costes marginales.



2.3 Leyes aplicadas al Open Data: Normas Técnicas de Interoperabilidad (NTI)

- Las NTI fueron propuestas por el Esquema Nacional de Interoperabilidad (ENI)
- Determinan una serie de pautas para publicar y reutilizar datos.
- Dotan a la estructura de datos de un buen formato para su reutilización.
- Interoperabilidad: capacidad de posibilitar el intercambio de datos



2.3 Leyes aplicadas al Open Data: NTI de reutilización

- Es una de las NTI más relevantes.
- Fue añadida por el Real Decreto 1495/2011.
- Su objetivo es establecer el conjunto de pautas básicas para la reutilización de los documentos y recursos elaborados por el sector público.
- Asegurar la persistencia de la información.
- Las variaciones que aportan son:
 - Esquema de identificadores (URIs)
 - Uso de metadatos en documentos y catálogos.
 - Taxonomía de sectores primarios
 - URIs correspondientes a los recursos geográficos del territorio español.

3. Protocolos y software para portales open data

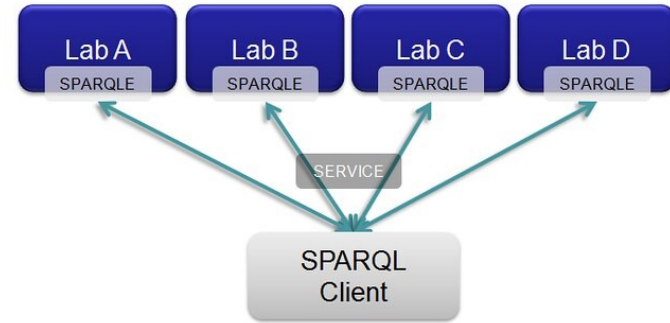
- 3.1 Open Data Protocol: Odata
 - Protocolo de código abierto basado en REST, HTTP y JSON.
 - Permite obtener y filtrar datos.
 - Fue una iniciativa impulsada por Microsoft.
 - Dispone de diversas librerías en .NET, Java, JavaScript, PHP, Python, C++, etc.
 - Su mayor ventaja es permitir añadir opciones de manipulación sobre las consultas realizadas

3. Protocolos y software para portales open data

- 3.2 SPARQL Protocol and RDF Query Language: SPARQL

- Realiza consultas sobre grafos RDF.
- Gran potencial para aplicaciones OpenData.
- Peticiones GET con una query.

Federated SPARQL Query



SPARQL Client sends subqueries into each endpoint and merges the results

3. Protocolos y software para portales open data

- 3.3 DCAT-AP

- Vocabulario RDF → consultas sobre catálogos.
- Ayuda a la descripción de catálogos.
- Catálogo: “conjunto de datos en el que cada elemento individual es un registro de metadatos que describe algún recurso”.
- Es la aplicación de DCAT a portales de datos.
- Objetivo: aumento de la interoperabilidad.

3. Protocolos y software para portales open data

- 3.4 Software de apoyo para portales: características
 - Apoyo para el desarrollo de portales open data.
 - La finalidad debe de ser la de ofrecer un catálogo sobre los datos del portal.
 - Funciones:
 - Catálogo de datos.
 - Servicio de datos (acceso a los datos, actualización, visualización y galería de aplicaciones).
 - Servicios informativos.
 - Servicios para la comunidad.

3. Protocolos y software para portales open data

- 3.5 Software de apoyo para portales: CKAN
 - Creación y gestión de portales
 - Previsualización de datos con mapas, gráficos y tablas.
 - Open source.
 - Gran número de extensiones (datastore, multilenguaje, stats, data viewer, etc)
 - Posibilidad de implementaciones externas.



comprehensive knowledge
archive network

3. Protocolos y software para portales open data

- 3.6 Software de apoyo para portales: DKAN
 - CMS para datos abiertos (sistema de gestión de contenidos).
 - Basado en Drupal.
 - Open source.
 - Gran facilidad para desarrollar sitios webs.
 - Módulos para gestión de datos.

4.1 Formatos abiertos

- Formato abierto → especificaciones del software que dan el formato disponibles.
 - De forma gratuita.
 - Sin limitaciones en la reutilización.
 - Compatibilidad entre distintos software.
- TXT, RTF, PDF, Office Open XML, OpenDocument, HTML.

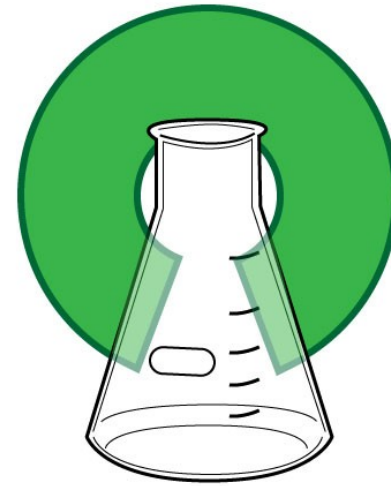
4.2 DOCX

- Características de DOCX
 - Microsoft word
 - Basado en XML
 - Guarda información en forma de XML y CSS.
 - Al guardar se comprime.
- Strict
 - Estandarización ISO/IEC 29500 en 2008
 - Ajustes de compatibilidad (espacios, paginación, márgenes, etc)
 - Agentes externos a Microsoft
- No-strict
 - Formato por defecto



4.3 Ciencia abierta

- Relación con los formatos abiertos
 - Publicación de cuadernos, datos, información recopilada, etc.
 - Elección de un formato adecuado que facilite su reutilización.



open science

4.4 Formatos en investigaciones

- A la hora de publicar documentos se recomienda:
 - Elegir un formato estándar
 - Documentos compatibles.
- Para elegir el formato:
 - Utilizar software no propietario.
 - Estándares abiertos
 - ASCCI, UNICODE, etc.
 - No encriptados y sin compresión.

