



Glossari de termes de Dades obertes

En aquest document es defineixen de manera senzilla alguns dels termes més habituals en el món de les Dades obertes. Les fonts d'informació emprades per a la seva redacció són diverses i les pots trobar al final d'aquest document.

Terme

Definició

Agent reutilitzador / infomediari

Tota persona, física o jurídica, que reutilitza informació del sector públic, ja sigui per a finalitats comercials o no comercials, sempre que aquest ús no constitueixi una activitat administrativa pública.

L'agent reutilitzador/infomediari administra informació en favor de tercers per mitjà de la recopilació de dades de diverses fonts, el seu estudi i selecció, per organitzar-les i redistribuir-les de manera rellevant per als seus usuaris.

Anàlisi de dades

Procés d'inspecció, depuració, transformació i modelització de les dades amb l'objectiu d'obtenir informació útil, proposant conclusions i afavorint la presa de decisions mitjançant múltiples aproximacions i tècniques en àrees com l'empresa, les ciències i les ciències socials.

Anonimització

Processament de les dades per garantir que cap individu no pugui ser identificat a partir de les dades que es publiquen. L'anonimització contribueix al compliment de la normativa en matèria de protecció de dades de caràcter personal. Les principals tècniques són l'agregació de registres d'individus i l'anonimització d'aquelles característiques que ajudarien a identificar les persones individualment.

API (Application Programming Interface)

En referència a les dades, és un canal proporcionat perquè programes o apps puguin llegir les dades obertes directament des d'una pàgina web.

L'app envia a l'API una consulta preguntant per determinades dades. Això permet a l'app utilitzar les dades sense necessitat de descarregar tot el conjunt de dades, garantint que aquestes sempre estan actualitzades.

Apps/ Aplicacions

Software informàtic dissenyat per ajudar els usuaris a desenvolupar tasques específiques utilitzant dispositius intel·ligents. En el món Open data, les aplicacions permeten maximitzar el potencial de les dades.

Atribució

Reconeixement de la font de dades quan s'utilitza, es reutilitza o es redistribueix. Les llicències d'ús poden requerir l'atribució a la font.

Big data

La complexitat d'aquestes dades fa que el seu emmagatzematge, transmissió o processament no es puguin fer per mitjans tradicionals. La necessitat creixent de processar aquests conjunts de dades ha donat lloc al desenvolupament de tecnologies informàtiques específiques, architectures i llenguatges de programació d'elevada sofisticació tècnica.

Catàleg de dades

Llista ordenada de conjunts de dades posats a disposició de tothom per al seu ús, reutilització i redistribució.

Terme

Definició

Conjunt de dades

Recull de dades relacionades, convenientment estructurades i organitzades, de manera que puguin ser tractades (processades) per obtenir-ne informació. Els conjunts de dades tenen forma de taula. Acostumen a estar disponibles per al seu accés o descàrrega en un o diversos formats. Estan habitualment allotjats en un repositori i poden pertànyer a un o més grups.

Exemple: conjunt de dades del padró municipal d'habitants d'un territori en un període determinat.

Expressió mínima de contingut sobre un tema, que es pot representar de manera numèrica, alfabètica o algorítmica. Mostra un atribut o una característica.

Dada

Quan les dades s'estructuren d'una manera formalitzada i adequada per a la comunicació, interpretació o processament per mitjans automàtics i/o humans, es transformen en informació.

Dades del sector públic

La tasca de les administracions públiques comporta l'ús de grans volums de dades, moltes de les quals no són confidencials. El valor d'aquestes dades es pot maximitzar posant-les a disposició de tothom en formats oberts.

Són productes i serveis d'informació que han estat generats, creats, recopilats, processats, preservats, mantinguts, disseminats o finançats per un govern o una institució pública i que compleixen els requisits i les restriccions legals pertinents.

Depuració de dades

Processament d'un conjunt de dades per fer-lo intel·ligible per a l'ús, reutilització o redistribució. Pot comportar corregir inconsistències i errors, eliminar elements que no es poden llegir per ordinador, utilitzar terminologia harmonitzada, garantir la validesa i la fiabilitat dels nombres i altres quantitats o la conversió a formats apropiats. La depuració està molt relacionada amb la qualitat de les dades.

Document

Tota informació, en qualsevol suport (material o electrònic), o forma d'expressió (textual, gràfica, sonora o audiovisual) incloent les metadades associades.

Documents o recursos d'informació reutilitzables

Documents en mans de les administracions, òrgans i entitats del sector públic emprats per persones físiques o jurídiques, amb finalitats comercials o no comercials, sempre que aquest ús no constitueixi una activitat administrativa pública, d'acord amb l'àmbit d'aplicació i exclusions de la llei.

Federació de dades obertes

Acció de publicar les dades obertes d'un organisme en el Portal d'una altra institució d'abast més gran, amb la finalitat de divulgar les dades obertes a un públic més ampli.

Format

Descripció de com es representa un arxiu en un ordinador. El format normalment correspon a la part final del nom de l'arxiu (l'extensió). El format es refereix a l'estructura interna de l'arxiu, no a la manera com es mostra als usuaris. Per exemple, un arxiu CSV i un XLS s'estructuren de manera molt diferent, però tenen una aparença similar quan s'obren des d'un full de càlcul.

Format llegible per màquina

Format d'arxiu estructurat que permet a les aplicacions informàtiques identificar, reconèixer i extreure dades específiques amb facilitat.

El material no digitalitzat no és llegible per màquina. Igualment, alguns materials digitals, com els arxius PDF tampoc no són llegibles per màquina pel fet que no és senzill extreure dades amb facilitat.

Alguns exemples de formats llegibles són **CSV**, **JSON** o **XML**.

Gestió de dades

Les polítiques, els procediments i les decisions tècniques emprades relacionades amb el cicle de vida de les dades, des de la recollida, passant per l'emmagatzematge, la preservació i l'ús.

Terme

Definició

Grup, col·lecció, categoria

S'utilitzen per gestionar i agrupar dades que tenen característiques en comú, com ara el tema que tracten. Per exemple: economia, població o eleccions.

Infografia

La categorització dels catàlegs de dades es recull en la normativa. En el cas espanyol, a la Norma Tècnica d'Interoperabilitat de Reutilització de recursos de la informació.

Representació visual de dades amb un objectiu informatiu. Poden incloure signes lingüístics i no lingüístics formant descripcions, seqüències expositives, argumentatives o narratives. Les infografies són útils per explicar informació de manera didàctica.

Informació

Un seguit de dades estructurades amb una disposició que la gent pot entendre i processar. La informació es converteix en coneixement quan es contextualitza.

Integració de dades

Les dades més interessants són aquelles que combinen diferents fonts. Per aconseguir-ho, és necessari que els conjunts de dades siguin compatibles entre ells: això implica que cal emprar els mateixos noms per als mateixos objectes. Quan la qualitat de les dades és alta, la integració és més senzilla.

Interoperabilitat

Propietat o capacitat de dos o més sistemes per intercanviar informació i utilitzar-la. Un dels objectius del moviment de les Dades obertes és afavorir la interoperabilitat entre les bases de dades de les administracions públiques.

Linked data (Dades enllaçades)

Les Dades enllaçades és la forma que té la web semàntica de vincular diverses dades que estan distribuïdes al web, de manera que es referencien de la mateixa manera que ho fan els enllaços de les pàgines web.

Metadades

Dades associades a un conjunt de dades que el descriuen i proporcionen informació addicional sobre aquest. És informació estructurada que descriu, explica o localitza un conjunt de dades i en facilita l'ús, la recuperació i l'administració.

Algunes de les metadades més habituals són el títol, la data de publicació, l'autor, la descripció dels continguts, les etiquetes, les llicències per a la reutilització o el responsable del conjunt de dades.

L'harmonització de les metadades entre organitzacions és essencial per facilitar l'accés, la cerca de dades, la reutilització i la comparació entre conjunts diferents.

Mineria de dades

Conjunt de tècniques estadístiques avançades que s'empren per al descobriment i la obtenció d'informació existent de forma implícita en les dades, i que és útil per a un determinat àmbit d'estudi.

Obertura de dades (Open Data)

Acció i resultat de fer accessibles dades públiques de les organitzacions, mostrant-les de manera que es puguin reaprofitar per qualsevol agent reutilitzador. Més concretament, una dada és oberta si qualsevol és lliure d'utilitzar-la, reutilitzar-la i redistribuir-la amb l'únic límit del requisit d'atribució de la font i/o reconeixement d'autoria.

Portal de Dades obertes

Plataforma web per publicar dades. Frequentment s'utilitza com a sinònim de catàleg de dades, si bé el portal proveeix serveis que complementen la interfície de cerca d'un catàleg.

És més precís definir-lo com un software que permet abordar els aspectes essencials de les dades obertes com ara la publicació de dades, la gestió de les metadades, l'emmagatzematge, el control d'accés, les cerques i les visualitzacions. Les seves potencialitats s'incrementen quan inclouen visualitzacions interactives de les dades.

Qualitat de les dades

El conjunt de dades ideal és precís, complet, permanentment actualitzat, consistent en els termes emprats i en el tractament dels casos perduts, llegible per màquina, segueix la nomenclatura habitual en el seu camp d'estudi, i es publica amb les metadades necessàries perquè l'usuari en conegui al detall les característiques.

Terme

Definició

Repositori de dades

Un servei en línia d'emmagatzematge de dades sense mecanisme de cerca.

Reutilització de la informació del sector públic

Ús de documents en mans de les administracions i organismes del sector públic, per part de persones físiques o jurídiques, amb finalitats comercials o no comercials, sempre que aquest ús no constitueixi una activitat administrativa pública. Queda exclòs d'aquest concepte l'intercanvi de documents entre administracions i organismes del sector públic en l'exercici de les funcions públiques que tenen atribuïdes.

Sistemes d'Informació Geogràfica (SIG)

Les eines geoespaciales són sistemes que integren, emmagatzemen, editen, analitzen, comparteixen i mostren informació geogràfica per a la presa de decisions. Les aplicacions SIG permeten a l'usuari crear consultes interactives, analitzar informació espacial, editar dades en mapes i presentar els resultats.

Visualització

Representació gràfica de dades que ajuda l'usuari a comprendre l'estructura i el significat de la informació continguda en elles. El seu potencial per a l'anàlisi s'incrementa quan són interactives, és a dir, quan permeten que l'usuari personalitzi allò que vol analitzar.

Web semàntica

Infraestructura de tecnologies i mecanismes que ofereix la possibilitat de definir, integrar, compartir i reutilitzar informació al web entre diferents parts de forma automatitzada en funció del seu significat.

Fonts d'informació:

- Glossari de l'Open Data Monitor
- Glossari de la Fundación CTIC
- Glossari de l'Open Data Handbook
- Glossari del Project Open Data
- Glossari del Portal de Dades obertes de la ciutat de Nova York
- Glossari del Portal de Dades obertes del Govern dels Estats Units d'Amèrica
- Glossari del Portal de Dades obertes del Govern del Regne Unit
- Glossari de Termes clau OCDE (Enquesta Open Data Government)
- Glossari de la Guia estratègica de dades obertes de la Federació Espanyola de Municipis i Províncies
- Norma Tècnica d'Interoperabilitat de Reutilització de Recursos d'Informació
- RD 1495/2011, del 24 d'octubre, pel qual es desenvolupa la llei 37/2007, del 16 de novembre, sobre reutilització de la informació del sector públic
- Llei 37/2007, del 16 de novembre, sobre reutilització de la informació del sector públic, modificada per la Llei 18/2015, del 9 de juliol
- Guia d'aplicació del RD 1495/2011 pel qual es desenvolupa la llei 37/2007, del 16 de novembre, sobre reutilització de la informació del sector públic
- Portal de Dades obertes del Govern d'Espanya
- Portal de Dades obertes de la Generalitat de Catalunya
- Portal de Dades obertes de la Diputació de Barcelona
- Portal de Dades obertes de l'Ajuntament de Barcelona