



G CONSELLERIA  
O PRESIDÈNCIA  
I I ADMINISTRACIONS  
B PÚBLIQUES  
/ ESCOLA BALEAR  
ADMINISTRACIÓ PÚBLICA

<b>UNITAT 1</b>	<b>INTRODUCCIÓ A L'OBERTURA I LA REUTILITZACIÓ DE DADES PÚBLIQUES</b>
<b>CONTINGUTS</b>	<a href="#">1. Introducció: definicions i conceptes bàsics.....2</a> <a href="#">2. Principis generals de les dades obertes.....5</a> <a href="#">3. La qualitat de la dada.....5</a> <a href="#">4. Beneficis de les dades obertes.....12</a> <a href="#">5. Problemes associats a l'obertura de dades.....15</a>
<b>OBJECTIUS</b>	<ol style="list-style-type: none"><li><b>1. Conèixer les definicions i conceptes bàsics: dada oberta, <i>dataset</i>, metadada, format llegible per màquina, reutilitzador, catàleg, portal, etc.</b></li><li><b>2. Comprendre els principis generals de les dades obertes: accessibilitat, gratuïtat, granularitat, no discriminatòries</b></li><li><b>3. Entendre la importància de la qualitat de la dada: actualització</b></li><li><b>4. Tenir clars els beneficis de les dades obertes</b></li><li><b>5. Conèixer els problemes associats a l'obertura pública de dades</b></li></ol>

Autor/a: Servei de Sistemes d'Informació

Data d'elaboració: 20/03/2023

Aquesta obra es difon mitjançant la llicència [Creative Commons Reconeixement-NoComercial-CompartirIgual 4.0 Internacional](#).

## 1 Introducció: definicions i conceptes bàsics

Avui en dia, en un món cada vegada més digitalitzat i globalitzat, les dades constitueixen un element clau per a empreses i organitzacions, les quals fonamenten la presa de decisions, el disseny d'estratègies i la implementació de canvis a partir de l'anàlisi d'informació. L'objectiu és, doncs, revertir la presa de decisions a cegues o basades en la simple intuïció i basar-la en anàlisis exhaustives de dades, que permetin conèixer la situació actual, preveure el que pot ocórrer en un futur pròxim o mesurar el rendiment de les accions realitzades.

En aquest context, les administracions públiques són conscients del potencial que té la informació que generen o recopilen en l'exercici de les seves funcions i, en conseqüència, de la importància que suposa l'impuls de polítiques d'obertura de dades públiques. Aquesta obertura comporta la publicació de dades confiables, pertinents i oportunes, fet que hauria de permetre a totes les persones i organitzacions que vulguin emprar-les generar valor sobre aquesta informació.

Així doncs, podem definir les dades obertes com aquelles dades que poden ser utilitzades, reutilitzades i distribuïdes per qualsevol persona, amb l'objectiu de tractar-les o generar-ne de noves en l'àmbit públic i privat, en condicions de lliure accés, sense limitacions de caràcter tècnic o legal (exceptuant el manteniment de la privacitat en dades personals) i publicades en formats oberts i estructurats.

### Definicions prèvies

Seguidament, presentam un conjunt de definicions prèvies que ens ajudaran a descriure els conceptes que manejarem durant el curs:

- **Dada.** Representació de fets, conceptes o instruccions de manera formalitzada i adequada per a la comunicació, la interpretació o el processament per mitjans automàtics o humans. Exemple: la xifra d'aturats en una província durant un determinat any.
- **Conjunt de dades.** Dades relacionades, convenientment estructurades i organitzades, de manera que puguin ser tractades (processades) apropiadament per obtenir-ne informació. Exemple: historial d'aturats per província.
- **Distribució.** Informació en un format concret, accessible des d'un URL concret. Un recurs d'informació pot disposar d'una o de múltiples distribucions. Exemple: un fitxer descarregable en CSV o una API de consulta en línia per a la mateixa informació.
- **Metadades.** Dades que defineixen i descriuen les dades. És informació estructurada que descriu, explica o localitza un conjunt de dades, cosa que fa més senzilla la recuperació d'informació, la utilització o l'administració d'un conjunt de

dades. Exemple: una fitxa que contengui la descripció del conjunt de dades, incloent-hi la data de publicació, l'autor i els possibles valors que poden prendre el camp, la província o el període.

- **Dades obertes.** Dades que es poden utilitzar, reutilitzar i redistribuir lliurement, amb l'únic límit, si escau, del requisit d'atribució de la font o el reconeixement de l'autoria.
- **Informació del sector públic.** Productes i serveis d'informació que han estat generats, creats, recol·lectats, processats, preservats, mantinguts, disseminats o finançats per un govern o una institució pública, i que compleixen els requisits i les restriccions legals pertinents.
- **Reutilització de la informació del sector públic (RISP).** Ús de la informació que està en mans de les administracions i els organismes del sector públic. Aquest ús és duit a terme per persones físiques o jurídiques, amb finalitats comercials o no comercials.
- **Agent reutilitzador o infomediari.** Tota persona, física o jurídica, que reutilitza informació del sector públic, ja sigui per a finalitats comercials o per a finalitats no comercials, sempre que aquest ús no constitueixi una activitat administrativa pública.
- **Repositori de dades.** Servei en línia d'emmagatzematge o *hosting* sense mecanisme de recerca. Pot ser tan senzill com un servidor web amb una carpeta i fitxers estàtics, sense un índex addicional o categorització, a excepció d'una pàgina principal per a cada conjunt de dades. Exemple: una carpeta publicada a través d'un servidor web on figuren fitxers descarregables en format CSV.
- **Catàleg de dades.** Col·lecció de metadades rellevants sobre conjunts de dades. A diferència del repositori de dades, el catàleg de dades se centra en l'organització dels conjunts de dades, mentre que el repositori es refereix al sistema d'emmagatzematge. El catàleg, normalment, no té en compte on estan emmagatzemades les dades. Podrien estar publicades en el mateix servidor web, juntament amb el catàleg (és a dir, en aquest cas, el catàleg contindria el repositori de dades). També podrien estar distribuïdes a través del web, amb un catàleg que simplement apuntàs a aquestes adreces remotes. En aquest cas, ens referiríem al catàleg com un agregador de dades o un indexador de dades. Exemple: catàleg de dades *datos.gob.es*.
- **Portal de dades obertes.** Web que complementa la interfície del catàleg. S'hi pot oferir informació sobre els formats en què es publiquen les dades, notícies relacionades, formularis de contacte, etc. Sovint, des d'aquests portals es pot accedir als catàlegs de dades.
- **Sector infomediari.** Conjunt de totes les empreses que fan servir les dades obertes per crear nous serveis i productes per oferir a la societat.

## Què són les dades obertes?

Les dades obertes (*open data*, en anglès) són les dades produïdes o recopilades per una administració pública o privada que es posen a l'abast de la societat perquè puguin ser emprades lliurement de manera senzilla i còmoda.

Les dades obertes són essencials per a la transparència de les administracions públiques (principi de govern obert) i disposen, per si mateixes, d'un gran potencial. Per aquest motiu, governs de tot el món impulsen polítiques de dades obertes, ja que aquestes dades es constitueixen com:

- El **nou or** sobre el qual se sustenta un nou model d'economia digital.
- El **nou substrat** sobre el qual es construeix un nou model de relació entre l'Administració i la ciutadania. Aquest nou model permet als ciutadans conèixer amb més detall l'activitat dels seus governs i administracions, i afavoreix la seva participació en la presa de decisions de forma molt més directa.
- La matèria primera per a l'**impuls de la innovació i el desenvolupament** social i econòmic.

### Com es posen a l'abast de la societat?

Les dades obertes són publicades mitjançant conjunts de dades (en anglès, *datasets*). Els *datasets* representen les dades de manera estructurada i en un format reutilitzable (CSV, JSON, XLSX). Cal notar, doncs, que el format PDF no és un format reutilitzable.

Els *datasets* es defineixen i s'identifiquen mitjançant metadades (identificador, nom, descripció, data de creació i tipologia).

### Com es carreguen les dades obertes en un *dataset*?

La càrrega de dades en un *dataset* es pot fer de manera manual o automàtica:

- De manera **manual**: es fa mitjançant fitxers de dades estàtics. Cal que hi hagi una persona encarregada d'anar carregant els fitxers per mantenir actualitzades les dades del *dataset*. La tasca d'actualització és més costosa i es corre el perill que el *dataset* quedi desactualitzat.
- De manera **automàtica**: es fa mitjançant API (*application programming interface*). Són funcions que extreuen automàticament les dades de la base de dades i les carreguen al *dataset*. Cal definir una freqüència d'actualització perquè aquestes funcions, de manera programada, extreguin periòdicament les dades i mantinguin actualitzat el *dataset*. Facilita el manteniment actualitzat del *dataset* amb poc esforç.

### Com la societat pot emprar o consumir les dades obertes d'un *dataset*?

Qui estigui interessat en les dades d'un *dataset* les pot consumir de dues maneres:

- Descarregant les dades del *dataset* en un fitxer estàtic. El format d'aquest fitxer sempre serà reutilitzable: full de càlcul (XLSX), CSV, JSON, etc.
- Accedint a les dades mitjançant API. De la mateixa manera que un *dataset* es pot carregar de manera automàtica mitjançant una funció d'API, també és possible consumir automàticament les dades d'un *dataset* per mitjà d'aquestes funcions.

Quan s'empren les dades obertes, s'ha de fer referència a la font o al reconeixement de l'autoria.

## 2 Principis generals de les dades obertes

Per assegurar que les dades obertes són fàcilment reutilitzables, es requereix el compliment d'una sèrie de principis bàsics:

- **Completes.** Totes les dades públiques han d'estar disponibles, sempre que no estiguin subjectes a restriccions de privadesa o seguretat.
- **Primàries.** Les dades han d'estar disponibles en les fonts primàries, sense haver estat modificades o processades de cap manera.
- **Periòdiques.** Les dades han d'estar disponibles tan aviat com sigui possible per preservar-ne el valor i mantenir-les actualitzades de la mateixa forma.
- **Accessibles.** Les dades han d'estar disponibles de manera que qualsevol persona sense discriminació pugui emprar-les, modificar-les i compartir-les per a qualsevol finalitat.
- **Processables.** Les dades disponibles han d'estar estructurades de manera que permetin ser processades de forma automatitzada.
- **No discriminatòries.** Les dades han d'estar disponibles per a qualsevol persona sense necessitat de cap tipus de registre.
- **No propietàries.** Les dades han d'estar disponibles en un format que no sigui d'ús exclusiu de cap entitat (no propietàries) per evitar restriccions d'ús.
- **De llicència lliure.** Les dades públiques no han d'estar subjectes a drets d'autor, llicències o patents.

## 3 La qualitat de la dada

L'obertura de les dades, des del punt de vista normatiu i tècnic, no és suficient per crear un ecosistema de reutilització de dades pròsper: en cas que fallin la disponibilitat i la qualitat de les dades, això podria tenir efectes negatius no només en la reutilització de les dades, sinó també en la credibilitat de les institucions que les publiquen.

La qualitat de les dades és una de les principals traves per al sector reutilitzador, i és un aspecte fonamental per garantir-ne la reutilització i donar suport a l'ecosistema de la creació de serveis i aplicacions. La baixa qualitat de les dades obertes obstaculitza una reutilització eficient d'aquestes dades a causa de la necessitat d'invertir recursos, per part dels usuaris, en la comprensió dels *datasets* mal documentats i per la realització addicional de tasques de depuració i processament.

Per incentivar la reutilització de les dades obertes, cal que les administracions inverteixin en la millora de la seva qualitat. S'han de definir un conjunt de directrius per millorar la qualitat de dades.

La disponibilitat de *datasets* per a la reutilització es pot aconseguir mitjançant la descàrrega d'arxius o accedint a serveis de dades (API). Els arxius de dades descarregables constitueixen la materialització completa d'un *dataset*. En canvi, l'accés a dades mitjançant serveis permet obtenir múltiples *datasets* en funció de les consultes que el reutilitzador configuri. Igualment, s'anomena *distribució* la forma en la qual es representen els *datasets* (és a dir, els formats). Aquests són diversos i responen a les necessitats del reutilitzador: pot ser un full de càlcul, CSV o Excel, un arxiu XML, un arxiu d'imatge, dades vinculades en RDF, etc. La disponibilitat de formats alternatius de cada *dataset* facilita la reutilització.

Dos aspectes importants que contribueixen a millorar la qualitat general són l'estandardització, l'enriquiment i la documentació de dades obertes.

### **Què és la qualitat de les dades obertes?**

Hi ha múltiples definicions de *qualitat de les dades*. Amb tot, la més adequada seria la següent: la idoneïtat d'un *dataset* per donar resposta al propòsit per al qual ha estat creat. Aquesta definició implica que la qualitat de les dades ha de ser gestionada sobre el propòsit que té el *dataset*. La gestió de la qualitat de les dades, a més, ha de proporcionar mètodes i eines per avaluar i establir processos de millora sempre que siguin precisos.

Al llarg d'aquests últims anys, la majoria d'iniciatives de dades obertes apliquen mètodes i eines per avaluar i establir processos de millora de la qualitat de les seves dades. Moltes d'aquestes iniciatives han emprat el mètode de cinc estrelles per a les dades obertes, publicat per Tim Berners-Lee el 2006. Tot i que aquest esquema és àmpliament utilitzat per mesurar la qualitat de les dades, només cobreix un aspecte específic de la qualitat: la codificació emprada per publicar les dades. Per aquest motiu, un conjunt de dades publicat pot aconseguir el nivell de cinc estrelles, però, al mateix temps, pot mostrar una qualitat deficient, ja que pot presentar altres tipus d'errors, com ara errors de sintaxi, duplicitat de registres o dades obsoletes.

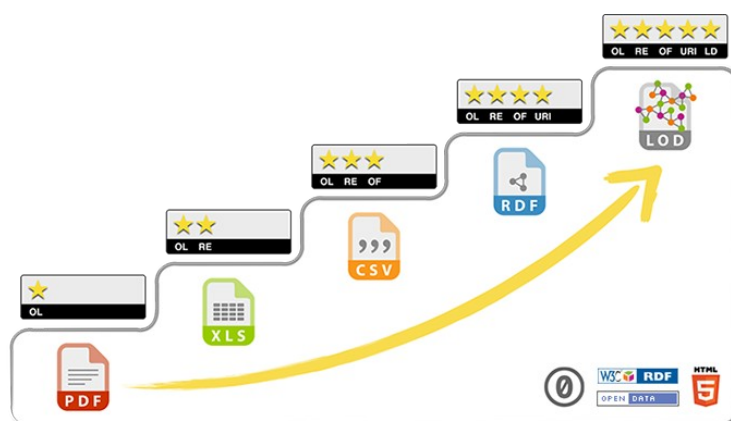


Figura 1: Mètode de 5 estrelles per a les dades obertes publicat per Tim Berners-Lee

Igualment, és comú identificar iniciatives que lliguen la qualitat de les dades amb els principis que han de regir tota política d'obertura de dades. Aquests principis ens indiquen que les dades han de ser completes, primàries, actuals, accessibles, processables per màquines, no discriminatòries, no propietàries i sense restriccions d'utilització. Amb tot, aquests principis —tot i que hi guarden relació— no posen el focus en la qualitat de les dades, sinó en una sèrie de propietats que han de posseir els conjunts de dades per considerar-los oberts i reutilitzables.

En la literatura hi ha diferents enfocaments sobre quines característiques han de tenir les dades de qualitat, i això comporta que existeixin diferents models de mesurament i assegurament.

A continuació, se citen una sèrie de referències que estableixen característiques imprescindibles per aconseguir una alta qualitat en les dades:

- Els principis definits per la Carta Internacional de les Dades Obertes (Open Data Charter) estableixen ja una sèrie de qualitats que les dades de qualitat han de complir, com ara completesa, exhaustivitat, puntualitat, oportunitat, comparabilitat i interoperabilitat. No obstant això, hi ha altres aspectes que també defineixen la qualitat de les dades i que han de tenir-se en compte a l'hora de produir i valorar qualsevol tipus de dades.
- L'OCDE (Organització de Cooperació i Desenvolupament Econòmic), en el seu informe *Marc de qualitat i directrius per a les activitats estadístiques de l'OCDE*, publicat el 2011, considera la qualitat en termes de set dimensions: pertinència, precisió, credibilitat, actualitat, accessibilitat, interpretabilitat i coherència. A més, estableix que la qualitat està àmpliament vinculada amb les perspectives, les necessitats i les prioritats dels reutilitzadors.
- D'altra banda, la Norma ISO/IEC 25012 també estableix un model de qualitat del producte de dades compost per quinze característiques, classificades en dues grans categories. Dotze es relacionen directament amb les dades com a producte: exactitud, completesa, consistència, credibilitat, actualitat, accessibilitat, conformitat, confidencialitat, eficiència, precisió, traçabilitat i comprensibilitat.

Un altre enfocament diferent, tot i que en línia amb la qualitat que han de presentar les dades, és la proposta que el 2016 va publicar la revista *Scientific Data* de *Nature*, sobre els «principis FAIR per a la gestió i administració de dades científiques». Els principis FAIR són un conjunt de directrius precises i mesurables que ha de seguir qualsevol científic per a la publicació de les seves dades amb la màxima qualitat possible. Aquests principis posen l'accent en la capacitat d'acció dels sistemes informàtics per trobar, accedir, interoperar i reutilitzar les dades amb la mínima intervenció humana, ja que els éssers humans depenen cada vegada més del suport informàtic per tractar les dades com a resultat de l'augment del volum, la complexitat i la velocitat de creació de les dades. Vegem quins són aquests principis:

- **Trobables (*findable*):** les dades i metadades han de poder ser trobades per la comunitat de manera senzilla després de ser publicades, mitjançant eines de recerca.
  - Han de tenir assignat un identificador únic i persistent.
  - Cal descriure les dades amb metadades de manera adient.
  - S'han d'indexar les dades i metadades en el recurs de recerca.
  - En les metadades s'ha d'especificar l'identificador de les dades que es descriuen.
- **Accessibles:** les dades i metadades han de ser accessibles, i per això han de poder ser descarregades per altres investigadors mitjançant els seus identificadors.
  - Les dades i metadades han de ser recuperables mitjançant el seu identificador, a partir de protocols de comunicació estandarditzats.
  - El protocol ha de ser obert, gratuït i d'aplicació universal.
  - El protocol ha de permetre un procediment d'autenticació i autorització quan sigui necessari.
  - Les metadades han de ser accessibles, fins i tot quan les dades ja no estiguin disponibles.
- **Interoperables:** les dades han de poder integrar-se amb altres dades. A més, les dades han de poder interoperar amb aplicacions o fluxos de treball per poder ser analitzades, emmagatzemades i processades.
  - Les dades i metadades han d'emprar un llenguatge formal, accessible, compartit i àmpliament aplicable.
  - Les dades i metadades han d'emprar vocabularis.
  - Les dades i metadades han d'incloure referències qualificades a altres dades o metadades.
- **Reutilitzables (*reusable*):** les dades i metadades han d'estar ben descrites perquè puguin ser reproduïdes o combinades en diferents entorns.



- Les dades i metadades han d'estar ben descrites amb una pluralitat d'atributs precisos i rellevants.
- Les dades i metadades han d'estar associades a una procedència detallada.
- Les dades i metadades han de complir les normes de la comunitat pertinents per al sector.

### Característiques de qualitat de dades

Els referents esmentats anteriorment defineixen un ampli conjunt de característiques comunes que s'espera que compleixin els *datasets* considerats d'alta qualitat.

A continuació, es descriu succintament cadascuna d'aquestes característiques:

- **Exactitud/precisió.** Tot i que no són termes equivalents, es refereixen a la veracitat que proporcionen les dades. Les dades que presenten aquesta característica representen correctament el valor veritable de l'atribut mitjançant el qual simbolitzen el món real, a més dels mesuraments que són precisos, consistents i reproduïbles.
- **Completesa.** Les dades es consideren completes quan tota la informació requerida per a un atribut està disponible. Les dades han de presentar un nivell de detall i una disgregació adequada per ser rellevants i reutilitzables.
- **Consistència/coherència.** Les dades han d'estar lliures de contradiccions i tenir coherència lògica en un context específic, per exemple, de format o temporal.
- **Credibilitat,** tant per a les dades en si com per a la font d'informació. Les dades han de ser objectives, han d'estar publicades amb els estàndards estadístics apropiats, i les pràctiques i polítiques per a la seva recollida i publicació han de ser transparents. La credibilitat també inclou el concepte d'autenticitat (la veracitat dels orígens de dades, atribucions i compromisos).
- **Actualitat i actualització / puntualitat.** Les dades han d'estar disponibles a temps i sense retards que afectin la seva rellevància, i s'han d'actualitzar regularment, a fi de mantenir-ne el valor.
- **Accessibilitat,** referida a la facilitat d'accés a les dades. Les dades han d'estar disponibles per a la més àmplia gamma d'usuaris i propòsits.
- **Conformitat.** Les dades s'han d'adherir a estàndards o normatives vigents.
- **Confidencialitat.** Les dades s'han de publicar respectant-ne la privadesa i la seguretat. En contextos específics, les dades només poden ser accessibles i interpretades per usuaris autoritzats. La confidencialitat és un aspecte fonamental de la seguretat de la informació.
- **Eficiència.** Les dades tenen atributs que poden ser processats i proporcionats amb uns recursos raonables.

- **Traçabilitat** respecte a la font o l'origen de les dades. Les dades tenen atributs que proporcionen un historial del camí d'accés auditat a les dades o qualsevol altre canvi realitzat sobre elles.
- **Comprensibilitat/interpretabilitat.** Les dades poden ser interpretades i llegides pels usuaris i ser expressades emprant llenguatges, símbols i unitats coherents amb el context. Certa informació sobre la comprensibilitat pot ser expressada mitjançant metadades.

## Per què és important que les dades disposin de la màxima qualitat possible?

Generalment, la solució dels problemes de qualitat de les dades en conjunts de dades obertes implica una inversió significativa, de vegades amb rendiments decreixents, tant per part dels publicadors com per part dels usuaris.

Ja l'OCDE, en el *Marc de qualitat i directrius per a les activitats estadístiques de l'OCDE*, esmentat més amunt, determina que la rendibilitat (costs i beneficis) és un factor que s'ha de tenir en compte en qualsevol anàlisi de qualitat, ja que pot afectar totes les característiques abans esmentades.

El 2011, Haug *et al.* van dur a terme una anàlisi de les causes i una estimació dels costos associats a la qualitat deficient de les dades. Els autors afirmen que, a la pràctica, la baixa qualitat de les dades pot implicar perjudicis econòmics a l'organització que publica les dades de múltiples formes. A més, determinen que les organitzacions tendeixen a sobreestimar la qualitat de les seves dades i a infravalorar el cost generat pels errors derivats. En nombroses ocasions, els publicadors de dades gasten més diners a solucionar errors derivats de dades deficientes que a evitar amb antelació els potencials problemes que el seu ús a llarg termini pot generar en els reutilitzadors.

Aquesta anàlisi sosté que el nivell òptim de manteniment de qualitat de les dades no és aconseguir dades perfectes, sinó assolir un equilibri entre el cost de les tasques associades amb l'assegurament de la qualitat de les dades i l'estalvi del cost causat per dades de baixa qualitat.

A més, els autors classifiquen els tipus de costos com:

- Costos causats per la qualitat deficient de les dades, ocasionada per:
  - Costos directes associats a la verificació, la reentrada i la compensació de dades.
  - Costos indirectes derivats de decisions o accions incorrectes i pèrdua d'inversions.
- Costos en millora o assegurament de la qualitat de les dades, derivats de:
  - Costos de prevenció associats a la formació, el monitoratge, el desenvolupament i la implementació estàndard.
  - Costos de detecció associats a la creació d'anàlisis i informes relacionats amb les dades.
  - Costos de reparació associats a la planificació de reparacions i la reparació d'implementacions.

En el concepte de les dades obertes és important tenir en compte que els costos derivats de la mala qualitat de les dades transcendeixen en la repercussió que té sobre la mateixa organització que les gestiona i que les empra en primera instància, atès que es traslladen al sector reutilitzador, de manera que es multiplica l'efecte negatiu que això comporta.

A més a més, és essencial que els conjunts de dades estiguin acompanyats d'unes metadades de qualitat excel·lent, ja que és el primer contacte que l'usuari té amb el conjunt de dades. Per a una reutilització efectiva de les dades, és imprescindible que estiguin acompanyades d'una documentació que ajudi els reutilitzadors a comprendre el contingut dels *datasets* i a entendre com s'han d'emprar, inclosos els defectes ja detectats.

Les metadades formen part del *dataset* i proporcionen informació crucial sobre l'origen de les dades, el grau d'actualització, les restriccions legals sobre el seu ús i altra informació rellevant. A més, les metadades haurien d'informar sobre la qualitat que presenten les dades. Unes metadades deficientes disminueixen la probabilitat de reutilització de les dades.

### **Pautes generals per garantir la qualitat de les dades obertes**

La qualitat insuficient de les dades afecta negativament el procés de reutilització d'aquesta informació. Això dificulta l'objectiu prioritari que persegueixen les iniciatives de dades obertes, que és aconseguir la màxima reutilització i generació de valor a partir de dades públiques.

A continuació, es detallen algunes de les causes dels problemes de qualitat de dades més freqüents:

- Utilitzar formats de dades no processables.
- Emprar codificació de caràcters no estandarditzats.
- Anomenar inadequadament les columnes.
- Publicar dades incompletes o amb valors absents.
- Incloure registres duplicats.
- No estandarditzar els valors de les dades.
- Proporcionar una quantitat inadequada de dades per facilitar-ne l'anàlisi.
- Emprar un format inadequat de variables de data i hora.
- Utilitzar un format inadequat de tipus de dades numèriques.
- Barrejar escales numèriques.
- Mesclar rangs en un mateix conjunt de dades.
- No incorporar variables amb informació geogràfica.
- Incorporar subtotals, totals o agrupaments.
- Presentar dificultats per accedir-hi.
- Organitzar inadequadament els *datasets* disponibles.

## 4 Beneficis de les dades obertes

Les dades que estan en mans de les administracions públiques tenen un **gran potencial** per a la generació de beneficis econòmics, socials i per a la mateixa Administració:

- **Econòmics.** Diversos estudis estimen que el volum de negoci entorn de les dades els anys vinents serà molt elevat.
- **Socials.** L'obertura de dades també genera una altra sèrie de valors per a la ciutadania: la reutilització de les dades ofereix als ciutadans la possibilitat de millorar la seva informació i el seu coneixement sobre les activitats del sector públic, i permet fer-les més transparents i properes.
- **Administracions.** Per a les administracions públiques, la reutilització de la informació del sector públic permet revertir en la societat els beneficis en termes de coneixement i econòmics, i contribueix a incrementar la seva transparència.
- **Generació de valor.** Perquè les dades obertes arribin a produir beneficis per a l'Administració, es necessita tota una estratègia activa que impliqui els diferents agents de l'ecosistema de la reutilització.

### Beneficis econòmics

Ja són diversos els estudis econòmics, tant a escala nacional com a escala europea o internacional, que estimen que el volum de negoci entorn de les dades els anys vinents serà molt elevat. El darrer estudi de caracterització del sector infomediari elaborat per l'Observatori Nacional de les Telecomunicacions i de la Societat de la Informació (ONTSI) estima:<sup>1</sup>

- Un volum de negoci de les empreses infomediàries espanyoles d'entre 450 i 500 milions d'euros l'any 2014.
- La generació d'entre 4.200 i 4.700 ocupacions a Espanya, majoritàriament llocs de treball qualificats i de qualitat.

### Quina és l'estimació de volum de negoci que generaran les dades obertes a escala mundial?

Tant a escala europea com a escala internacional, i també en el marc del G20, s'han dut a terme una sèrie d'anàlisis per sectors clau, i s'han generat una sèrie d'informes.

#### EUROPA

En el conjunt d'Europa, l'estimació estaria entorn dels 200.000 milions d'euros de beneficis fins a l'any 2020, si es combina el potencial de les dades obertes amb el del *big data* en un conjunt de 21 sectors estratègics. Això suposaria un

<sup>1</sup> Font: *Estudi de caracterització del sector infomediari a Espanya. Part I. Anàlisi del sector infomediari públic* (2014, ONTSI).

increment de l'1,9 % del PIB del conjunt de la Unió, que en termes actuals és més d'un any complet de creixement econòmic.

## **INTERNACIONAL**

L'anàlisi internacional també mostra un potencial de gairebé quatre bilions i mig d'euros anuals tan sols per a un total de set sectors clau: educació, transports, productes de consum, electricitat, indústries extractives, sanitat i finances personals.

## **G20**

En la mateixa línia, es pronostica un creixement acumulat entorn dels deu bilions d'euros per al conjunt de les economies del G20 durant els pròxims cinc anys. Això suposaria una mica més del 50 % de l'objectiu de creixement establert per a aquest mateix període.

## **Beneficis socials**

L'obertura de dades genera altres valors per als ciutadans:

- Més informació. La reutilització de les dades ofereix la possibilitat de millorar la informació i el coneixement de la ciutadania sobre les activitats del sector públic.
- Més transparència. La millora de la informació i el coneixement sobre el sector públic genera més transparència i proximitat a la ciutadania.
- Creació de nous productes i serveis que cobreixen les necessitats de la ciutadania.
- Creació de nous llocs de feina i generació de valor econòmic.

Aquests nous productes i serveis es poden crear a partir de les dades obertes:

- Àmbit de la informació meteorològica: definició de prediccions meteorològiques personalitzades.
- Àmbit de la informació jurídica: recopilacions de lleis i de jurisprudència internacional i nacional.
- Àmbit de la informació cartogràfica: sistemes de navegació per a vehicles i informació cartogràfica digital.
- Àmbit de la investigació: ampli accés a les dades provinents de recerques científiques que ajuden els investigadors de diferents àmbits a col·laborar en el mateix conjunt de dades, a emprendre formes completament noves de recerca científica, a explorar les correlacions entre els resultats de les investigacions o a crear condicions més adequades per abordar els desafiaments mediambientals.

- Altres fonts d'informació: informació sobre subvencions, estadística, turisme, estudis econòmics o de trànsit.

### **Beneficis per a les administracions**

Per a les administracions públiques, la reutilització de la informació del sector públic produeix els beneficis següents:

- Revertir en la societat els beneficis derivats de la reutilització de la informació que generen i gestionen en l'exercici de la seva activitat. Fomenten nous serveis per a la ciutadania i noves oportunitats de negoci en el sector de la reutilització de les dades.
- Contribuir a incrementar la seva transparència.
- Fomentar el dret al coneixement, principi bàsic de la democràcia, pel fet de donar publicitat a procediments polítics, judicials, econòmics i administratius.
- Definir una guia per a la participació democràtica. Fomenta el desenvolupament de projectes col·laboratius, en què la ciutadania s'uneix per aportar solucions innovadores a problemes socials, o bé contribueix a la millora dels serveis públics existents.
- Acostar l'Administració pública a la ciutadania i a les empreses (i eliminar les barreres geogràfiques).
- Contribuir a revalorar l'activitat pública.
- Contribuir a millorar la qualitat de les dades i el monitoratge i l'ajustament de polítiques públiques.
- Compartir i emprar dades d'altres departaments de la mateixa Administració.
- Augmentar el coneixement de les dades que gestionen i dels procediments administratius que duen a terme.
- Augmentar la interoperabilitat entre administracions.
- Tenir un coneixement més profund del seu mateix negoci (punts forts i mancances), pel fet de permetre l'ús i l'anàlisi de les dades.

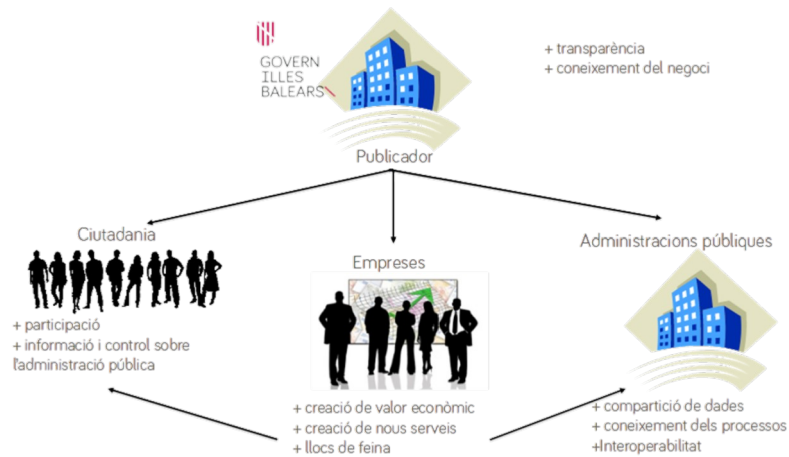


Figura 2: Beneficis per a les Administracions públiques

## 5 Problemes associats a l'obertura de dades

**Amb quines dificultats ens trobam a l'hora de dur a terme l'obertura de les dades?**

- L'obertura de les dades no es percep com una prioritat.
- Es pensa que les dades d'un departament o servei no tenen interès per a la ciutadania i les empreses.
- Hi ha un sentiment de propietat de les dades públiques.
- Hi ha por cap a l'obertura de les dades (es tem que es pugui vulnerar la LOPD o el secret estadístic).
- Es tendeix a pensar que les dades d'un departament o servei no són útils per a altres serveis de la mateixa Administració o per a altres administracions públiques.

A continuació, es presenten les barreres més rellevants amb què ens trobam a l'hora d'obrir les dades:

- **Culturals.** Hi ha una falta de conscienciació de les organitzacions públiques sobre el potencial que posseeixen les dades obertes, així com un gran temor a la pèrdua de control d'aquesta informació. Canviar la mentalitat de les administracions requereix:
  - Un sòlid compromís polític en el grau més alt per canviar la mentalitat de les administracions.



- Un diàleg dinàmic entre els agents interessats, incloses les administracions, les empreses i la comunitat acadèmica.
- Una formació i una sensibilització dels treballadors públics.
- **Legals i organitzatives.** Encara hi ha una sèrie de buits legals i d'aspecte organitzatiu que dificulten la creació de nous productes i serveis. També s'han de tenir en compte la protecció de dades de caràcter personal i els drets de la propietat intel·lectual com a fets que fan difícil l'obertura de les dades. Segons el sector infomediari, les comunitats autònomes publiquen les dades de manera poc homogènia i això en dificulta la reutilització i la interoperabilitat. Amb quins altres problemes es troba el sector infomediari?
  - Les administracions públiques no saben quines són les obligacions de la Llei de reutilització.
  - Hi ha una falta de criteri a l'hora d'oferir dades.
  - Hi ha una falta d'usabilitat de les dades.
  - Hi ha una falta d'estandardització dels formats de les dades.
  - Hi ha una falta d'interès de les administracions públiques per conèixer les necessitats de les empreses.
  - Hi ha una falta de resposta de l'Administració quan les empreses fan sol·licituds d'obertura de dades.
- **Pràctiques i tècniques.** La diversitat tecnològica pot ser una barrera a l'hora de tenir dades interoperables. Les dades són interoperables si podem combinar les dades de diferents organitzacions i països. Per fomentar la interoperabilitat, a l'Estat espanyol s'ha desenvolupat la Norma tècnica d'interoperabilitat de reutilització de recursos de la informació. El fet que els sistemes d'informació de les administracions no estiguin preparats per publicar mitjançant dades obertes per defecte és una altra barrera a l'obertura de les dades, de la mateixa manera que també ho és la falta d'un estàndard semàntic únic de dades.

## BIBLIOGRAFIA

Kroes, Neelie (2011). «Data is the new gold», Press Conference on Open Data Strategy.

Llei 18/2015, de 9 de juliol, per la qual es modifica la Llei 37/2007, de 16 de novembre, sobre reutilització de la informació del sector públic.

Llei 37/2007, de 16 de novembre, sobre reutilització de la informació del sector públic.

McCandless, David. «Data Is the New Soil», TED Talk on Visualizing Data.

Ministeri d'Afers Econòmics i Transformació Digital (2022). *Guía práctica para la mejora de la calidad de datos abiertos*.

Open Government Working Group (2007). *The Annotated 8 Principles of Open Government Data*.

Reial decret 1495/2011, de 24 d'octubre, pel qual es desenvolupa la Llei 37/2007, de 16 de novembre, sobre reutilització de la informació del sector públic, per a l'àmbit del sector públic estatal.

Reial decret 4/2010, de 8 de gener, pel qual es regula l'Esquema Nacional d'Interoperabilitat en l'àmbit de l'administració electrònica.

Resolució de 19 de febrer de 2013, de la Secretaria d'Estat d'Administracions Públiques, per la qual s'aprova la Norma tècnica d'interoperabilitat de reutilització de recursos d'informació.

Reutilització de la informació del sector públic (RISP).

Ubalde, B. (2013). «Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives», *OECD Working Papers on Public Governance*, núm. 22, OECD Publishing.

W3C. *Guía breve de linked data*.