

EAPC - Gestió de la informació: transparència i protecció de dades



La qualitat: factor d'èxit dels projectes de dades

22-01-2024

Ha plogut una mica des que les administracions públiques van impulsar projectes per permetre l'accés a la informació pública a reutilitzadors potencials. La missió d'aleshores ens és ben familiar: transparència, empoderament de la ciutadania en la supervisió de la gestió pública, transferència de coneixement i creació de valor acadèmic, econòmic i social. L'avenç imparable de la tecnologia en les darreres dècades i el canvi de paradigma de la transformació digital que va precipitar la situació d'emergència per la pandèmia de la COVID-19, han transformat l'escenari i actualment les administracions públiques disposen de portals de dades obertes on podem consultar informació de tot tipus i de qualsevol àmbit.

Certament, el catàleg de dades obertes de la Unió Europea compta amb més d'1.600.000 conjunts de dades procedents dels portals dels diferents països i regions; entre aquests, trobem el [Portal de dades obertes de Catalunya](#), amb més de 1.100 conjunts de dades, o el del [Consorci d'Administració Oberta de Catalunya \(AOC\)](#), amb més de 670 conjunts de dades dels diferents ajuntaments de Catalunya, entre d'altres.

Aquestes xifres podrien fer pensar que som davant d'un projecte madur, ja consolidat i amb un full de ruta amb pocs reptes per davant. Tanmateix, si revisem la normativa publicada els darrers anys, i tenim en compte el creixement dels volums de dades impulsat per la revolució de l'Internet de les coses (IdC), o la innovació disruptiva que ha precipitat la intel·ligència artificial, podem afirmar que les dades obertes segueixen sent un element crucial per proveir empreses, governs i organitzacions d'informació de qualitat sobre la qual prendre decisions, per definir estratègies, impulsar la innovació o millorar resultats.

És a dir, com a part de l'ecosistema de dades, les dades obertes són un element crucial en l'[Estratègia de dades de la Unió Europea \(2020\)](#), que té per objectiu crear un mercat únic de dades que en permeti, com el cas de les mercaderies, la lliure circulació dins de la UE i entre diferents sectors, per al benefici d'empreses, investigadors i administracions públiques. I és que la Comissió Europea estima que l'economia de la dada pot assolir el 2025 un valor de 829.000 milions d'euros, un 4% del PIB, i ocuparà gairebé 11 milions de professionals.

En aquest mercat únic els principis de les dades obertes segueixen sent vigents; de fet, els marcs legals i de referència que cerquen desenvolupar aquesta economia de la dada, o els mateixos sistemes d'intel·ligència artificial, requereixen de dades segures, de qualitat.

És el cas d'una de les normes de referència dins de l'Estratègia europea, el [Reglament de governança de dades](#), que estableix la definició de polítiques, procediments i estàndards per a la gestió integral de les dades. De fet, amb el Reglament com a referència, la mateixa Generalitat de Catalunya impulsa i comença a implantar el seu propi model de governança el 2023, amb l'aprovació del [Model de govern de les dades](#).

Tant en el cas de la normativa europea com en la de la Generalitat de Catalunya, la prioritat de la qualitat de les dades dins del marc de la governança assegura que les dades, independentment del seu cicle de vida, estiguin ben gestionades i proporcionin un valor òptim a l'organització. Precisament aquesta gestió, que ha de derivar en unes dades de qualitat, evitarà la fragmentació i la pèrdua d'informació o de coherència quan s'operi entre sistemes d'informació i plataformes diferents.

La interoperabilitat és encara més necessària en els [espais de dades](#), un entorn en què diversos actors –organitzacions i empreses– comparteixen informació de forma voluntària, segura, sobirana i interoperable. En aquests espais, cada participant manté el control sobre les seves dades, i per tant les normes de qualitat s'han de seguir de forma escrupolosa per garantir-ne una compartició eficaç.

L'Estratègia europea de dades considera la creació d'espais de dades en deu camps estratègics i d'interès públic (salut, agricultura, indústria, energia, mobilitat, finances, administració pública, qualificacions) i el seu desenvolupament s'ha de dur a terme tenint en compte elements com la millora de la posada a disposició de les dades, i la seva qualitat i interoperabilitat, entre d'altres.

Precisament un dels elements fonamentals per al desenvolupament d'aquests Espais de dades són les dades d'alt valor. Es tracta de conjunts de dades que aporten un gran valor afegit i importants beneficis a la societat. La Comissió Europea va identificar ([Reglament d'execució \(UE\) 2023/138 de la Comissió, de 21 de desembre de 2022](#)) sis àmbits temàtics: geoespacial, observació de la Terra i medi ambient, meteorologia, estadística, societats i propietat de societats, i mobilitat; d'altra banda, a finals de 2023 es va fer una proposta per ampliar la llista de categories, afegint-hi: pèrdua climàtica, energia, finances, administració pública i govern, salut, justícia i llengua. A més d'identificar els àmbits, el Reglament reflecteix les modalitats de publicació i reutilització, i fa una especial referència a la difusió de dades a través d'interfícies de programació d'aplicacions ([API\(Application Programming Interface\)](#)).

Per tant, tant la normativa de dades d'alt valor, la de reutilització d'informació del sector públic, com els fonaments dels espais de dades o les darreres especificacions UNE sobre dades, fan referència a la qualitat de les dades.

La qualitat

Hi ha diferents estàndards, normativa o polítiques que fan referència a la qualitat i la completesa de les dades. Cal que les definicions de cada característica o dimensió i els paràmetres de mesura d'una dimensió o característica en concret siguin compartits per tothom qui treballi en dades, de forma que s'asseguri una actuació homogènia sobre cada conjunt de dades. Per exemple, el [5star](#)

[Open Data](#) de Tim Berners-Lee recomana publicar les dades sota una llicència oberta, tan estructurades i granulades com sigui possible (com més detallades millor), utilitzant formats no propietaris (CSV enlloc d'Excel), relacionant dades.

Segons els [principis FAIR](#), els conjunts de dades haurien de complir un conjunt de qualitats precises i mesurables perquè siguin fàcils de localitzar, accessibles, interoperables i reutilitzables. No obstant això, aquests són aspectes específics relacionats amb la naturalesa del disseny de les dades, els quals no garanteixen necessàriament que les dades puguin utilitzar-se si no tenen el nivell de qualitat adequat.

Per contra, hi ha diferents especificacions UNE que introdueixen bones practiques pel que fa al Govern de les dades i la gestió de la qualitat de les dades: [UNE 0077](#), [UNE 0078](#) i [UNE 0079](#).

La norma UNE 0079:2023 de gestió de qualitat de les dades obertes parteix de les 15 propietats de qualitat especificades a la [norma ISO/IEC 25012](#): exactitud, completesa, consistència, fiabilitat, actualitat, accessibilitat, conformitat, confidencialitat, eficiència, precisió, traçabilitat, comprensibilitat, disponibilitat, portabilitat i recuperabilitat.

D'altra banda, els errors més freqüents en la qualitat de les dades obertes són els següents:

- Dades incompletes o limitades.
- Dades fragmentades (sèries temporals incompletes, en diferents arxius).
- Manca de detall.
- Dades sense actualitzar.
- Metadades actualitzades per a dades desactualitzades.
- Formats no reutilitzables o no estandarditzats.
- Falta de claredat en les llicències.

Pel que fa al contingut:

- Incloure diferents valors en la mateixa columna (per exemple, separats per comes, incorporant-hi en una mateixa columna codi i denominació o enunciat).
- Sèries temporals amb barreja de rangs numèrics diferents.
- Codificacions de caràcters no estandarditzada.
- Codificació no estandarditzada: per exemple, nom de països, municipis sense apòstrof; format d'hora no estàndard.
- Ús dels valors nuls: s'hauria d'indicar o només quan sigui 0. Si no, cal deixar la cel·la buida.
- Utilitzar un llenguatge excessivament tècnic o sense possibilitat de documentació de referència.
- No anonimitzar o pseudoanonimitzar les dades.

Pel que fa a l'estructura dels conjunts de dades, cal evitar:

- Diferents fulls en un únic fitxer.
- Files de totals o subtotals.
- Més d'una fila de capçaleres (la capçalera ha d'estar a la primera fila).

Pel que fa a les metadades, cal garantir la documentació dels aspectes rellevants, com l'estructura, les definicions i les relacions, i garantir-ne l'actualització de forma periòdica per reflectir-ne l'estat.

Igualment, cal establir procediments per avaluar i mantenir la qualitat de les dades, i implementar controls de qualitat per identificar i corregir errors, o desenvolupar polítiques per garantir la seguretat de les dades i la seva protecció contra amenaces, i assegurar-se que les

pràctiques de gestió de dades es fan d'acord amb les normatives de privacitat i les lleis aplicables.

Elements clau

En l'era de la intel·ligència artificial, assegurar la representativitat, l'equitat i l'exactitud de les dades esdevé imperatiu per evitar prejudicis i garantir l'entrega de resultats de qualitat. Les consideracions ètiques, la bona governança i el compliment de normes associades amb la qualitat de les dades, juguen un paper clau en la protecció contra prejudicis i desinformació i en el desenvolupament de l'economia de la dada.



Leticia Rodríguez i Riera

Cap de l'Àrea de Dades Obertes