

## BREVE GUIDA AGLI OPEN DATA

I contenuti informativi e i dati in possesso delle amministrazioni pubbliche, grazie alla continua evoluzione tecnologica, rappresentano una straordinaria occasione per **rendere più trasparenti le azioni dei governi e delle amministrazioni**, per **erogare servizi più efficienti** ma anche, favorendone il riuso da parte di altri soggetti pubblici o privati, per essere **utilizzati in ambiti differenti da quelli per i quali sono stati prodotti o raccolti**.

I dati pubblici, che nel passato erano destinati a rimanere relegati nell'ambito dei procedimenti amministrativi o di governo per i quali erano stati formati, oggi sono risorse sempre più accessibili, aperte e quindi conoscibili e fruibili da tutti.

### COSA SONO I DATI

Per spiegare meglio cosa sono gli Open Data, è necessario specificare cosa si intende per "dati".

I dati sono strettamente correlati a fatti, sono potenzialmente riproducibili, sono parte di strutture informative più vaste, possono essere lavorati e archiviati in formati digitali.

In poche parole: **dati → informazioni → conoscenza → consapevolezza**

Il dato, di per sé neutro, diventa informazione quando viene creato, estratto, elaborato e utilizzato per determinati obiettivi.

L'apparato di informazioni, spesso costituito da più dati dello stesso tipo o di tipo diverso (l'insieme di dati si definisce "**dataset**"), diventa **conoscenza** quando viene interpretato attraverso **strumenti, applicazioni, metodi, incroci, indicatori** etc.

La conoscenza, infine, acquisisce il valore di **consapevolezza** – nel caso degli Open Data la si può senz'altro definire di tipo "collettivo", nel senso più ampio di "bene comune" – quando ad essa viene attribuita la funzione di **cambiamento e miglioramento della realtà** (i fatti) interpretata e "aperta" grazie ai dati.

### OPEN GOVERNMENT DATA

Abbiamo visto nella nostra pillola video che gli Open Government Data sono dati pubblici in formato aperto e accessibili a tutti i cittadini, senza licenze di copyright o altri limiti legali all'utilizzo, che permettono di:

- **Rendere l'amministrazione trasparente.** Il libero accesso a documenti, atti e informazioni della PA e sulle scelte politico-istituzionali correlate è un aspetto fondamentale per la trasparenza delle istituzioni

e per la stessa democrazia, perché i cittadini possono così verificare l'efficienza della Pubblica Amministrazione. Del resto, i cittadini SONO la Pubblica Amministrazione.

- **Rendere l'amministrazione aperta.** Distribuire i dati pubblici in un formato aperto e libero da restrizioni sia dal punto di vista dell'accesso che del riutilizzo, rappresenta il presupposto di base affinché possa svilupparsi non solo un vero e proprio processo di collaborazione tra le istituzioni e la comunità dei cittadini sulle scelte di governo, ma anche la rielaborazione in forma nuova e diversa dei dati messi a disposizione. Mediante strategie di apertura dei dati della Pubblica Amministrazione, i cittadini hanno l'opportunità di riutilizzare e integrare i dati messi loro a disposizione, fino a sviluppare servizi e applicazioni a vantaggio dell'intera comunità di utenti.

Gli Open Data, inoltre, devono essere...

- **completi:** devono cioè comprendere tutte le componenti che consentano di esportarli, utilizzarli online e offline, integrarli e aggregarli con altre risorse e diffonderli in rete;
- **tempestivi:** gli utenti devono essere messi in condizione di accedere e utilizzare i dati presenti in rete in modo rapido e immediato;
- **accessibili:** i dati devono essere resi disponibili al maggior numero possibile di utenti senza barriere all'utilizzo, preferibilmente senza il ricorso a piattaforme proprietarie. Devono essere inoltre resi disponibili senza alcuna sottoscrizione di contratto, pagamento, registrazione o richiesta;
- **leggibili da computer:** è necessario che i dati siano *machine-readable*, cioè processabili in automatico dal computer;
- **in formati non proprietari:** i dati devono essere codificati in formati aperti e pubblici, sui quali non vi siano entità (aziende od organizzazioni) che ne abbiano il controllo esclusivo;
- **liberi da licenze che ne limitino l'uso:** i dati aperti devono essere caratterizzati da licenze che non ne limitino l'uso, la diffusione o la redistribuzione;
- **riutilizzabili:** gli utenti devono essere messi in condizione di riutilizzarli e integrarli, fino a creare nuove risorse, applicazioni e servizi di pubblica utilità;
- **ricercabili:** i dati devono essere facilmente identificabili in rete, grazie a cataloghi e archivi facilmente indicizzabili dai motori di ricerca.

#### **RAW DATA NOW!**

Ovviamente, dal punto di vista degli Open Government Data, è auspicabile che l'informazione venga rilasciata in formati di file aperti *machine-readable*. Tuttavia, qualora la scelta fosse tra il pubblicare dati non completamente rispondenti ai requisiti indicati o non pubblicarli affatto, la logica dell'Open Data propende verso la prima soluzione. In questo caso si parla di **"Raw Data Now"**: nella peggiore delle ipotesi (disporre di dati non aperti) è comunque preferibile distribuirli, anche se – appunto – in formato "raw", cioè "grezzo". L'auspicio è che se i dati pubblicati sono sufficientemente interessanti, la comunità degli utenti si preoccuperà di convertirli poi efficacemente in dati aperti (si parla, in questo caso, di **"data scraping"**, ossia letteralmente "grattare i dati").

## FORMATO APERTO E INTEROPERABILITA'

Per mettere a disposizione del pubblico i dati è possibile utilizzare una grande varietà di formati. Non tutti, tuttavia, rispondono ai requisiti necessari a definire tali dati “aperti”.

Il formato in cui le informazioni sono pubblicate, ossia la base digitale in cui sono memorizzate le informazioni, può essere infatti aperto o chiuso. Un **formato aperto** è quello in cui le specifiche per il software sono a disposizione di chiunque, gratuitamente, in modo che chiunque possa utilizzarle nel proprio software senza alcuna limitazione di riutilizzo imposto attraverso diritti di proprietà intellettuale. Quando il formato è chiuso, invece, può significare o che il formato è proprietario e le caratteristiche tecniche non sono pubblicamente disponibili, o che il formato del file è proprietario e, anche se le specifiche tecniche sono pubbliche, il riutilizzo è limitato.

La ragione fondamentale per cui è importante chiarire il significato di “aperto” e del perché utilizzare proprio questa definizione, può essere identificata in un termine: **interoperabilità**. L'interoperabilità è la capacità di diversi sistemi e organizzazioni di lavorare insieme. Nel nostro caso, è la **capacità di combinare una base di dati (database) con altre**. L'interoperabilità è la chiave per realizzare il principale vantaggio pratico dell'apertura: aumenta in modo esponenziale la possibilità di combinare diverse basi di dati, e quindi sviluppare nuovi e migliori prodotti e servizi.

Per di più, il vantaggio dei formati di file aperti è evidente, in quanto permettono agli sviluppatori di produrre software e servizi utilizzando tali formati. Questo riduce al minimo gli ostacoli per riutilizzare le informazioni in essi contenute. Usare formati proprietari, invece, può creare dipendenza dal software di terzi o dai detentori della licenza del formato. Nel caso peggiore, ciò può significare che l'informazione può essere letta soltanto utilizzando un certo pacchetto software, che potrebbe avere un costo proibitivo o divenire obsoleto nel corso del tempo.

## DAI DATI GREZZI (RAW DATA) AI LINKED OPEN DATA (LOD): FORMATI E TIPOLOGIE DI DATI

È utile fare adesso una carrellata dei principali formati, codici e contenitori di file utilizzati per la creazione, l'archivio e il rilascio di dati.

### CSV (Comma Separated Values)

Formato di file per insiemi di dati a base testuale, che rende particolarmente agevole l'importazione da o l'esportazione verso fogli di calcolo e database. Il formato CSV separa i valori attraverso virgole (“comma”). I file CSV possono risultare molto utili perché compatti e quindi adatti a trasferire grandi insiemi di dati con la stessa struttura. Tuttavia, il formato è così spartano che i dati, senza documentazione, sono spesso inutili, dal momento che diventa quasi impossibile identificare il significato delle diverse colonne. È pertanto fondamentale che sia fornita un'appropriata documentazione (**metadati**) insieme ai file in formato CSV.

### Fogli di calcolo

Molti soggetti posseggono informazioni archiviate in fogli di calcolo, come ad esempio Excel. Questi dati possono essere utilizzati assieme alla corretta descrizione di ciò che le colonne rappresentano. Tuttavia, in alcuni casi potrebbero esserci funzioni e formule nei fogli di calcolo scomodi da gestire.

### Database

I database consentono un accesso diretto ai dati. Hanno il vantaggio di consentire agli utenti di mettere insieme ed estrarre soltanto ciò a cui sono interessati. Vi sono tuttavia alcuni problemi di sicurezza nell'estrazione remota da un database. Peraltro, l'accesso al database risulta utile solo se la struttura del database e l'importanza delle singole tabelle e campi sono ben documentati.

### RDF (Resource Description Framework)

È un linguaggio che definisce in che modo le informazioni devono essere rappresentate online. RDF associa dati a informazioni e documenti che circolano in Rete, permettendo al computer di comprendere il contesto e il significato delle informazioni, consentendo inoltre l'interoperabilità tra più applicazioni che si scambiano informazioni su Internet. In altre parole, il formato RDF consente di rappresentare i dati in una forma che rende facilmente integrabili dati provenienti da fonti diverse.

### HTML (Hyper-Text Markup Language)

Linguaggio di marcatura che permette di descrivere la formattazione di un documento web, definendone la visualizzazione attraverso il protocollo HTTP. Il linguaggio HTML permette di generare link a documenti differenti, consentendo l'organizzazione e la costruzione di ipertesti. Oggi una gran quantità di dati è disponibile nel formato HTML su vari siti.

### XML (eXtensible Markup Language)

XML è un formato/linguaggio flessibile ampiamente utilizzato per lo scambio dei dati, perché fornisce buone opportunità per conservare la struttura dei dati, nonché consente agli sviluppatori di includere parte della documentazione insieme ai dati senza interferire con la loro lettura.

### JSON

JSON è un formato di file semplice e molto facile da leggere per tutti i linguaggi di programmazione. La sua semplicità risiede nella maggiore facilità di elaborazione per i computer rispetto ad altri formati.

### Documento di testo

Documenti in formati classici come Word o PDF possono essere sufficienti per mostrare alcuni tipi di dati. Possono essere condivisi a basso costo, essendo i formati in cui spesso i dati sono stati creati. Questi formati non aiutano però a mantenere coerente la struttura, e quindi è spesso difficile, se non impossibile, inserire i dati in modo automatico.

### Testo semplice

I documenti di testo (.txt) sono molto facili da leggere per i computer. Tuttavia, generalmente non includono informazioni strutturali all'interno del documento, con la conseguenza che gli sviluppatori devono creare un analizzatore in grado di interpretare ogni documento così come appare.

### Immagine acquisita (scanner)

Probabilmente la forma meno adatta per la maggior parte dei dati, anche se sia TIFF che JPEG sono in grado di aggiungere dettagli su ciò che è rappresentato nella foto (fino a inserire nell'immagine un contenuto con

testo integrale del documento). Possono tuttavia essere utili per la visualizzazione in immagini di quei dati che non sono stati originati elettronicamente.

### Formati proprietari

Alcuni sistemi specializzati hanno i propri formati di dati in cui possono salvare o esportare dati. A volte può anche essere sufficiente condividere i dati in questi formati, soprattutto se si prevede un uso successivo in un sistema simile a quello da cui provengono. Dovrebbe sempre essere indicato dove si possono trovare ulteriori informazioni su tali formati, ad esempio con un link al sito web del fornitore. In generale è consigliabile di condividere dati utilizzando formati non proprietari.

### **LE 5 STELLE DI TIM BERNERS-LEE**

Per distinguere i diversi formati utilizzabili nella codifica dei set di dati, l'inventore del Web, Tim Berners-Lee, ha proposto un [modello di catalogazione](#) che li classifica in base alle loro caratteristiche su una scala di valori da 1 (una stella) a 5 (cinque stelle):

★ Una Stella. È il livello base, costituito da file non strutturati: ad esempio un'immagine in formato grezzo (formati come .gif, .jpg, .png), un documento in formato Microsoft Word, un file in formato Adobe PDF. Una sola stella indica la semplice disponibilità di una informazione e di un dato on line, in un formato qualsiasi. I dati distribuiti in questo formato sono leggibili e stampabili dagli utenti, possono essere conservati su un PC e sono semplici da pubblicare. Tuttavia non sono un formato aperto in quanto non è possibile effettuare su di essi alcuna elaborazione.

★★ Due Stelle. Questo livello indica dati strutturati ma codificati con un formato proprietario. Ad esempio un documento in formato Microsoft Excel. I dati caratterizzati dalle due stelle non sono un formato aperto in quanto per elaborarli è necessario un software proprietario, tuttavia di norma possono essere convertiti – essendo dati strutturati – in dati aperti.

★★★ Tre Stelle. Questo livello indica dati strutturati e codificati in un formato non proprietario. Ad esempio il formato .csv (Comma Separated Values) al posto – ad esempio – del formato Microsoft Excel utilizzato nel caso precedente. È possibile effettuare elaborazioni sui dati senza esser costretti a utilizzare software proprietario. Si tratta del formato più semplice di dati aperti.

★★★★ Quattro Stelle. Questo livello indica dati strutturati e codificati in un formato non proprietario che li rende utilizzabili direttamente online. Pensate, ad esempio, a un dataset contenente gli indirizzi dei monumenti di una città opportunamente codificati: da qualsiasi software – o dal browser – è possibile collegarsi alla URL che indica il singolo monumento, potendolo ad esempio georeferenziare su una mappa.

★★★★★ Cinque Stelle. Questo livello indica quelli che vengono definiti **Linked Open Data (LOD)**, quei dati aperti, cioè, che – dal punto di vista del formato – oltre a rispondere alle caratteristiche indicate al punto precedente, presentano anche, nella struttura del dataset, collegamenti ad altri dataset. In altri termini, è possibile collegare dinamicamente tra loro più dataset, incrociando così informazioni provenienti da fonti diverse, eventualmente gestite da diverse Amministrazioni o Enti Privati.

## LE TIPOLOGIE DI INFORMAZIONI DEGLI OPEN DATA

Gli Open Data fanno di frequente riferimento a informazioni rappresentate in forma di database e riferite alla tematiche più disparate: cartografia, genetica, composti chimici, formule matematiche e scientifiche, dati medici e pratica, bioscienze, dati anagrafici, dati governativi, statistiche, economia etc.

Ad esempio, si possono distinguere alcune tipologie di dataset aperti secondo i seguenti insiemi tematici:



- **Geodati:** dati utilizzati per realizzare mappe, per esempio la localizzazione di strade e palazzi, la topografia, la visualizzazione dei confini, la georeferenziazione di esercizi commerciali etc.;
- **Cultura:** dati riferiti a opere e prodotti culturali (per esempio: titoli, autori etc.), e generalmente conservati da biblioteche, gallerie, archivi, musei;
- **Scienze:** dati prodotti come parte della ricerca scientifica, dall'astronomia alla zoologia;
- **Economia e Finanza:** dati relativi ai conti pubblici (entrate e spese), informazioni sui mercati finanziari (titoli, azioni, obbligazioni etc.);
- **Statistica:** dati prodotti da uffici e servizi statistici, indicatori sociali, economici, demografici etc.
- **Meteo:** i vari tipi di dati utilizzati per comprendere e prevedere il meteo e il clima;
- **Ambiente e Salute:** informazioni relativi all'ambiente (presenza e livello di fattori inquinanti, qualità delle acque, rifiuti.), ai tassi e cause di mortalità, all'incidenza di malattie in determinate zone etc.
- **Trasporti:** orari, percorsi, statistiche sui tempi di percorrenza etc.

## IL VALORE ECONOMICO DEGLI OPEN DATA

L'apertura dei dati pubblici non determina solo effetti importanti sulle amministrazioni pubbliche di molti Paesi nel mondo, attraverso la **semplificazione di processi** e il **risparmio dei tempi**, ma produce anche rilevanti impatti sull'economia complessiva: le stime internazionali segnalano vantaggi rilevanti, con effetti ampiamente misurabili anche sul sistema economico-imprenditoriale, per l'utilizzo di dati cartografici, dati sui trasporti, dati sui beni culturali, demografici, immobiliari, occupazionali e sociali per produrre applicazioni e servizi.

Quasi tutte le informazioni pubbliche hanno, del resto, un valore intrinseco, che va oltre l'assolvimento dei compiti istituzionali dell'ente che le detiene. Anche quando la singola informazione ha un'utilità limitata, spesso la lettura in serie storica, la raccolta organizzata e la combinazione di più informazioni con fonti differenti consente di riattribuire un valore rilevante a quella informazione, nonché generare un valore aggiunto notevole.

Così, ad esempio, le **informazioni geografiche** sono alla base di cartine e mappe, le quali possono servire a fornire servizi on line di ogni tipo, mentre i **dati immobiliari** possono servire a operatori e istituti di credito che vogliono fare incontrare la domanda e l'offerta di alloggi e per proporre soluzioni di finanziamento. I **dati meteorologici** possono essere utilizzati sia per servizi a valore aggiunto in tempo reale (ad esempio funzioni "push" per smartphone con previsioni selettive del tempo), sia per alimentare i siti meteo più tradizionali. I **dati sui trasporti** possono servire per alimentare servizi di informazioni sulla mobilità. I **dati sui bilanci delle imprese** possono offrire importanti informazioni di mercato per fare analisi della concorrenza e studi di scenario economico utili per i decisori pubblici o per chiunque debba realizzare strategia di investimento.

I dati e le informazioni pubbliche, se resi disponibili e fruibili, possono quindi trasformarsi in contenuti "attivatori" di nuove e talvolta imprevedibili attività e servizi a valore aggiunto per gli utenti costituiti da cittadini, imprese e amministrazioni.