

# Qualità dei dati aperti

Formez  PA



Questo materiale didattico è stato realizzato da Formez PA nel *Progetto PerformancePA*, Ambito A Linea 1, in convenzione con il Dipartimento della Funzione Pubblica, organismo intermedio del Programma Operativo Nazionale Governance e Azioni di Sistema (PON GAS), Asse E Capacità istituzionale. Il PON GAS è cofinanziato dal Fondo Sociale Europeo ed è a titolarità del Ministero del Lavoro e delle Politiche Sociali.

L'opera è distribuita con Licenza [Creative Commons Attribuzione - Condividi allo stesso modo 4.0 Internazionale](https://creativecommons.org/licenses/by-sa/4.0/).



*Autore:* Gianfranco Andriola

*Creatore:* Formez PA

*Diritti:* Dipartimento della Funzione Pubblica

*Data:* Ottobre 2015

## Qualità dei dati aperti

Questa videolezione è dedicata a comprendere insieme quali sono le caratteristiche dei dati aperti di qualità, cioè di cosa dovrà tener conto la Pubblica Amministrazione nel momento in cui realizza un'operazione di valorizzazione del proprio patrimonio informativo attraverso l'open data. Innanzitutto vedremo qual è la differenza tra un dato pubblico e un dato pubblico in formato aperto. Poi vedremo gli elementi che caratterizzano la qualità dei dati aperti (cioè singolarmente uno per uno) e quindi gli aspetti di cui bisogna tener conto nel momento in cui si pubblicano i dati; infine proveremo a fare una riflessione finale su quello che è il monitoraggio dell'attività di rilascio dei dati in formato aperto.

Non tutti i dati pubblicati dalla Pubblica amministrazione online possono essere considerati open data per farlo è necessario che abbiano tre caratteristiche principali cioè che siano: resi disponibili su internet; che siano resi disponibili in un formato che in gergo si dice "machine readable" cioè leggibile da altre applicazioni informatiche quindi un formato strutturato, quindi detto in maniera ancora più esplicita in un foglio di calcolo organizzato per celle e per colonne; infine che i dati siano rilasciati con una licenza aperta che garantisca a chiunque la possibilità di poterli riutilizzare. Infine, non tutti i dati rilasciati in formato aperto dalla Pubblica Amministrazione possono essere considerati open data di qualità ci sono degli elementi di cui bisogna tener conto per fare open data di qualità.

Non si deve pensare ai dati come numeri in maniera astratta. I dati nel momento in cui vengono esposti, nel momento in cui vengono rilasciati in formato open data vengono aggregati in file che in gergo vengono definiti "data set" cioè fogli di calcolo che contengono tutta una serie di dati. Vedremo insieme adesso come bisognerà tenere conto di alcune caratteristiche, di alcuni elementi per rendere questi data set di qualità. Innanzitutto non ci devono essere errori.

Le risorse contenute all'interno dei file devono essere descritte in maniera corretta; in questo caso si parla di accuratezza sintattica cioè tutte le risorse sono ben descritte e non ci sono refusi e non ci sono errori successivamente è necessario rilasciare i dati in maniera continuativa, in maniera sostenibile dati costantemente aggiornati. L'informazione, al di là dell'informazione pubblica, ma l'informazione in generale, è tanto più preziosa quanto più è aggiornata. Successivamente si può guardare in alcuni dati rilasciati dalla Pubblica Amministrazione la loro georeferenziazione. La georeferenziazione è estremamente importante ai fini del riuso. Molte applicazioni riutilizzano dati georeferenziati cioè dati che

possono essere collocati in punti precisi dello spazio e quindi bisognerà tener conto, per alcuni dati, di adottare formati e metodi di produzione che appunto tengano conto delle caratteristiche georeferenziali dei singoli dati.

Infine i dati devono essere completi cioè i numeri da soli significano praticamente nulla se non sono bene argomentati: il numero 12 che si vede in questa slide, da solo significa poco se non si sa esattamente di che cosa stiamo parlando, può essere 12 minuti 12 giorni 12 chilometri qualunque cosa che tenga il numero 12 come numero a disposizione quindi è necessario descrivere bene, argomentare bene in maniera completa i data set che vengono rilasciati.

Infine, i data set devono essere consistenti da un punto di vista interno, cioè non ci devono essere errori tra le risorse: se da un lato, nell'esempio, "Stefano risiede al Comune di Roma" nello stesso data set è impossibile che Stefano risieda nella Provincia di Como, c'è una contraddizione di termini interna; in questo caso si chiama appunto di consistenza interna al data set. Più o meno sulla falsariga di queste considerazioni si può fare anche una considerazione tra data set diversi: in questo caso si parla di consistenza esterna; ad esempio se da un lato Stefano del nostro esempio è sposato con Francesca, da un'altra parte quella stessa Francesca non può risultare nubile, quindi non ci devono essere contraddizioni in termini tra data set esterni l'uno all'altro altrimenti si cade in errore e si rischia di rendere più confusa l'operazione di apertura dei dati della Pubblica Amministrazione.

L'esposizione dei dati, la produzione dei dati, può essere vista esattamente come qualunque altro processo di produzione, cioè può essere organizzato per fasi, per fasi che ne prevedono la concretizzazione e poi delle azioni correttive per aggiustare il tiro. È molto comodo guardare alle tecniche di produzione già utilizzate in altri contesti come ad esempio il metodo PDCA che è l'acronimo di Plan, Do, Check ed Act. Di fatto si tratta di definire l'operazione, il processo di esposizione dei dati della Pubblica Amministrazione, per fasi quindi innanzitutto pianificarla (PLAN), cioè fare una cernita dei dati disponibili all'interno delle singole amministrazioni e vedere tra questi quali può essere più interessante iniziare a esporre in formato aperto; quindi (DO) quindi pubblicare i dati su internet, vedere quali sono le reazioni, vedere come la comunità che vuole riutilizzare quei dati reagisce, quindi un vero e proprio CHECK di monitoraggio, cioè capire come questi dati vengono riutilizzati e infine, se da questa attività di monitoraggio arrivano delle considerazioni interessanti su quelli che sono i dati più utilizzati piuttosto che i dati meno utilizzati, insomma, può essere un buon elemento per capire come sta andando e quindi proporre delle azioni correttive per riaggiustare il tiro e quindi incrementare per fasi successive il ciclo di produzione dei dati.

In questo caso è estremamente importante l'attività di monitoraggio anche perché è indubbio che per quanto l'open data sia una pratica per certi versi anche abbastanza economica comunque per la Pubblica Amministrazione può rappresentare un costo e ragionando in un'economia, cioè ragionando con risorse limitate, è necessario ed è sempre una buona norma tenere sotto controllo quella che è l'attività come sta andando, l'attività anche all'esterno. Quindi, il monitoraggio dell'uso dei dati e le analytics, tutto quello che ci permette di capire come, l'effetto che quei dati stanno avendo all'esterno della Pubblica Amministrazione è assolutamente indispensabile valutarlo in un'operazione di open data quindi monitorare i dati per porre delle azioni correttive che ne rendono ancora più focalizzato, ancora più interessante il riutilizzo e l'esposizione da parte della Pubblica Amministrazione.