

Governance e policy dei dati aperti



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: Vincenzo Patruno

Creatore: Formez PA

Diritti: Dipartimento della Funzione Pubblica

Data: Ottobre 2015

Una governance per gli open data

Fare open data per una pubblica amministrazione significa innanzi tutto **innovarsi** nella governance interna dell'ente. I dati sono un "effetto collaterale" delle attività di un determinato ufficio. Prendersi cura dei dati prodotti, prepararli e validarli per la loro pubblicazione sul catalogo open data vuol dire introdurre un cambiamento profondo nel modo di intendere l'attività quotidiana di ufficio e di servizio al cittadino. Significa valorizzare le competenze esistenti creandone allo stesso tempo di nuove. Riuscire a creare processi stabili di produzione e rilascio dati vuol dire migliorare l'organizzazione e l'efficienza interna degli interi processi amministrativi che generano quei dati, e dare opportunità di crescita ai dipendenti della pubblica amministrazione coinvolti, ai cittadini e alle aziende del territorio che possono beneficiare della disponibilità di dati di qualità. E' quello che chiamiamo **empowerment**¹ della pubblica amministrazione.

Un'esempio è fornito dal progetto Open Data Lazio. Il progetto infatti non ha voluto soltanto costruire un catalogo open data della Regione Lazio, ma ha puntato proprio sull'empowerment dei dipendenti al fine di migliorare l'attività amministrativa corrente e nello stesso tempo di costruire processi stabili per il rilascio di dati di qualità. E questo anche introducendo tra gli obiettivi individuali dei dirigenti la liberazione e pubblicazione dei dati facenti capo alle strutture da loro dirette.

Favorire il riuso dei dati

Per governare il processo di apertura dei dati è utile pensare all'essenza stessa dei dati aperti, ovvero il loro "riuso" per ulteriori scopi. Perché ciò avvenga, i dati devono rispettare una serie di caratteristiche sia di tipo legale, ad esempio la licenza d'uso associata al dataset², che di tipo più prettamente tecnologico, e cioè i formati con cui i dati vengono rilasciati.

In realtà, queste caratteristiche non sono sufficienti da sole a favorire il riuso di un dataset. Una ulteriore caratteristica che diventa determinante ai fini del riuso è quella legata alla qualità del dato. Un dato di bassa qualità è un dato che si presta molto poco ad essere riusato. Un dato di qualità è invece un dato che esprime tutte le proprie potenzialità e che si presta ad

¹ Nel processo di empowerment la pubblica amministrazione diventa consapevole del proprio potenziale e della possibilità di migliorare i propri processi interni, innovarsi e rinnovarsi.

² Collezione di dati, generalmente riguardanti una stessa organizzazione, che vengono esposti e gestiti congiuntamente.

essere elaborato, combinato e integrato con altri dati per generare nuova informazione e conoscenza, generando valore aggiunto.

Dataset di qualità

Come si può facilmente immaginare, la qualità del dato non è una caratteristica misurabile in modo deterministico. Non c'è in altre parole nessun "termometro" che possiamo utilizzare per dire che un dataset abbia una qualità uguale ad un certo valore. Possiamo però dare una serie di indicazioni che complessivamente fanno sì che il dataset pubblicato sia un dataset di qualità.

Una caratteristica di qualità è ad esempio la **completezza**. Un dataset è completo nel momento in cui viene pubblicato e diffuso assieme ai suoi metadati. I **metadati**³ sono delle informazioni fondamentali per far capire la struttura del dataset, il suo contenuto, come è stato ottenuto e quindi come utilizzarlo in modo corretto. Possiamo distinguere due tipi di metadati: mentre i metadati strutturali⁴ sono legati alla struttura stessa del dataset - è un metadato strutturale ad esempio la colonna "Comune" dell'elenco di tutte le farmacie d'Italia - i metadati descrittivi sono delle informazioni che vengono associate al dataset e che servono a capire come il dataset è stato definito. Un metadato descrittivo potrà essere ad esempio l'indicazione che i codici utilizzati per indicare i comuni sono i codici standard utilizzati da Istat.

Elementi che definiscono la qualità del dato sono poi la **tempestività** con cui il dato viene rilasciato, la **periodicità** con cui viene prodotto, il livello di dettaglio del dato. E' infatti fondamentale ai fini della qualità che il dato sia **primario**, ossia al massimo livello di granularità possibile, ovviamente senza che questo possa generare problemi legati alla confidenzialità e alla riservatezza del dato.

Processi di qualità

Un dato di qualità necessita di processi di qualità. Sono infatti i processi di qualità che possono generare dati di qualità. Intervenire sui processi interni nei quali vengono generati i dataset da pubblicare significa per una pubblica amministrazione intervenire in modo strutturale sulla produzione dei dati. Rendere stabile nel tempo la produzione di dataset di qualità e la loro pubblicazione sul catalogo dell'ente diventa quindi un elemento essenziale per far sì che i dati aperti possano generare interesse su chi poi i dati li utilizza.

³ Informazioni utili a descrivere un insieme di dati (dataset). I metadati risultano particolarmente utili per conoscere il contenuto di un dataset prima ancora di aprirlo, per aiutare i motori di ricerca a indicizzare al meglio le informazioni, a strutturare in maniera più funzionale possibile il catalogo che ospita i dataset.

⁴ Metadati che rappresentano la struttura del dataset e consentono di descrivere le relazioni esistenti tra le diverse risorse o tra parti di una stessa risorsa all'interno del dataset.

Quello che serve è quindi avviare all'interno degli uffici pubblici un "ciclo di vita" del dato che possa essere stabile e affidabile nel tempo.

Un team per gli open data

Come organizzarsi dunque per valorizzare il "ciclo di vita" dei dati aperti? L'AGID, l'Agenzia per l'Italia Digitale, nel pubblicare le *"Linee guida per la valorizzazione del patrimonio informativo pubblico"* ha preso molto in considerazione questi aspetti relativi ai processi e alla organizzazione interna. L'AGID suggerisce pertanto delle modalità operative che possano essere di riferimento per un ente pubblico nel momento in cui decide di organizzare la propria attività attorno agli open data. Un primo passo è quello di definire delle figure chiave all'interno dell'Ente pubblico. Va definito un "Team Open Data"⁵, un gruppo di lavoro con la responsabilità di gestire il processo di apertura dei dati.

All'interno del Team Open Data sarà poi opportuno identificare un Open Data Manager⁶, ossia la figura che, assieme a tutto il team ha la responsabilità di questo processo.

L'Open Data Manager collabora poi con altre figure: una di queste è il "Titolare della Banca Dati", generalmente un dirigente o un quadro responsabile del processo amministrativo che genera quel particolare dato.

Se prendiamo ad esempio l'ufficio SUAP, ossia lo sportello unico per le attività produttive di un comune, il "Titolare della Banca Dati" sarà il dirigente responsabile dell'ufficio SUAP.

A supportare il Titolare della Banca Dati vengono consigliate da AGID altre due figure: il "Responsabile tecnico della Banca dati", ossia colui che ha alcune conoscenze informatiche di base per poter operare sulla banca dati e il "Referente tematico della banca dati" che è un esperto di dominio e che conosce la storia dei dati prodotti da quel particolare ufficio.

Oltre agli uffici di produzione del dato, sono anche altri gli uffici che entrano in gioco: l'ufficio di statistica, che ha generalmente dimestichezza nell'utilizzare dati e che ha anche le competenze necessarie a validarli e strutturarli prima della loro pubblicazione, l'ufficio giuridico-amministrativo per le questioni legate agli aspetti legali del dato e l'ufficio comunicazione. Questo ovviamente dipende dalla grandezza e dalla complessità interna dell'ente pubblico in questione. Enti più piccoli potranno semplificare l'individuazione delle figure precedenti facendo ricadere ad esempio più funzioni su una singola figura.

⁵ Il Team Open Data "ha la responsabilità di pianificare e coordinare l'evoluzione continua dell'apertura dei dati nell'amministrazione, nonché dell'infrastruttura IT a supporto".

⁶ L'Open Data Manager "conosce i dati dell'amministrazione nel loro insieme, redige linee guida operative per lo scambio dati tra le diverse figure coinvolte e pianifica la strategia di apertura dei dati raccolti e analizzati e le attività di diffusione dei dati".

Come rilasciare dati aperti

Bene. Ora che il Team Open Data è composto, come prendersi cura dunque dei dati aperti e rilasciare dati di qualità?

Il punto di partenza è sicuramente una ricognizione interna dei dati prodotti dalle attività dei singoli uffici. Non va dimenticato che l'obiettivo non è quello di assemblare al volo un dataset da pubblicare in modo da incrementare il numero totale di dataset presenti sul catalogo, ma costruire un processo stabile che generi dataset di qualità. Ritorniamo all'ufficio SUAP del Comune e facciamo un esempio.

Partiamo dall'elenco delle attività produttive del comune e supponiamo di avere questo elenco all'interno di un foglio elettronico. Supponiamo di voler rilasciare il dataset con una frequenza annuale, con una data di riferimento prestabilita che può essere il primo gennaio di ogni anno. Il dataset sarà quindi una fotografia delle attività produttive di quel comune al primo gennaio di ogni anno. L'anno lo si potrà quindi inserire tra le colonne del dataset, diventando quello che abbiamo chiamato metadato strutturale. Il primo anno, il contenuto di questa colonna sarà identico per tutte le righe. L'anno successivo i nuovi dati non andranno a sostituire i primi, ma verranno aggiunti in coda al dataset dell'anno precedente. I nuovi dati saranno riconoscibili in quanto comparirà il valore del nuovo anno nella apposita colonna.

SUAP			
Attività Produttive			
XXX	XXX	XXX	Anno
XXX	XXX	XXX	2015
XXX	XXX	XXX	2015
XXX	XXX	XXX	2015
XXX	XXX	XXX	2016

Il dataset rilasciato sarà quindi un dataset il cui numero di righe crescerà di anno in anno. Così facendo viene preservata la storia del dato e l'integrità e coerenza del dataset. Questo aspetto diventa importante ad esempio per conoscere e quindi monitorare come cambia nel tempo la struttura produttiva del territorio. E' una cosa che può essere di interesse per le *policy* della pubblica amministrazione, per chi deve fare investimenti, per i cittadini che necessitano di accedere a servizi sul territorio, per il mercato immobiliare e così via.

Coinvolgere i cittadini attraverso l'open data

Abbiamo visto dunque che i dati hanno le potenzialità per generare un cambiamento all'interno delle nostre pubbliche amministrazioni. Pubblicare i dati prodotti da un particolare processo

amministrativo rende “visibile” quel processo innanzitutto alla stessa pubblica amministrazione, ma lo rende visibile anche agli *stakeholders* esterni⁷. A patto ovviamente che questi ultimi riescano a coglierlo. La cultura dei dati aperti va quindi promossa e sostenuta dalla pubblica amministrazione attraverso iniziative, incontri, patrocinii. Gli Open Data Day, che l’Istat ha patrocinato per le prime due edizioni, rappresentano un esempio del valore che i dati aperti possono avere per il nostro sistema Paese.

⁷ I dati possono apportare benefici e offrire opportunità a cittadini, community, ricercatori, giornalisti, aziende e startup.