



Unione europea
Fondo sociale europeo



*Ministero del Lavoro
e delle Politiche Sociali*
Direzione Generale per le Politiche
Attive e Passive del Lavoro



Progetto PerformancePA
Ambito A - Linea 1 - Una rete per la riforma della PA

Il mondo dei dati

Autore: Nello Iacono

Creatore: Formez PA, Progetto Performance PA

Diritti: Dipartimento della Funzione Pubblica

Ottobre 2015

Il mondo dei dati

Perché conoscere il mondo dei dati?

Conoscere il mondo dei dati diventa sempre più importante soprattutto per poter prendere decisioni “**informate**”, basate cioè su un quadro chiaro ed esauriente del contesto in cui si opera.

L'analisi dei dati è oggi resa più complessa per la grande quantità di dati disponibili e l'elevata digitalizzazione dei processi, che producono dati e si basano su dati.

Diventa quindi indispensabile comprendere alcuni aspetti fondamentali sui dati:

- **cosa sono;**
- **quali sono le fonti e le tipologie;**
- **come possono essere gestiti in termini di processi e di tecnologie.**

Comprendere i dati e il loro valore è **indispensabile** per capire:

- il **contesto interno** della propria organizzazione
- il **contesto esterno** del mercato e dell'utenza verso cui si rivolge.

Cosa intendiamo per dati

Cosa si intende qui per dato, e qual è il rapporto con il concetto di informazione e quindi di conoscenza?

- Un **dato** è un elemento informativo costituito da simboli che devono essere elaborati; l'elaborazione dei **dati** produce **informazioni**.
- Le **informazioni** sono quindi dati collocati in un contesto, che permette di associare loro un significato attraverso ad esempio, la categorizzazione, l'elaborazione, la correzione, la sintesi.
- Per **conoscenza** si deve intendere *"Un mix mutevole di esperienze strutturate, **informazioni** contestuali e intuizioni basate sull'esperienza, che forniscono un modello per valutare e incorporare nuove esperienze e informazioni."*, Davenport e Prusak(1998).

Qualità e affidabilità dei dati

Se i dati sono importanti, allora è fondamentale che siano affidabili e di elevata qualità.

La **qualità dei dati** non è un'espressione generica. Esiste infatti una norma internazionale ISO (ISO/IEC 2015) che definisce la qualità dei dati come combinazione di:

- **Qualità inerente**, data dalla capacità intrinseca di un insieme di dati di soddisfare i requisiti in un determinato contesto indipendentemente dall'ambiente tecnologico;
- **Qualità di sistema**, data dalla capacità di un insieme di dati di soddisfare i requisiti in un determinato contesto quando i dati sono memorizzati, usati o presentati in uno specifico ambiente tecnologico.

Per definire con maggior dettaglio qualità inerente e di sistema si sono definite 15 caratteristiche:

specifiche della qualità inerente

1. **accuratezza**: conformità ai valori reali (sintattica e semantica)
2. **aggiornamento**: conformità tempestiva ai valori reali (temporale); frequenza
3. **completezza**: presenza dei dati necessari (occorrenze, attributi, valori)
4. **consistenza**: assenza di contraddizioni interne
5. **credibilità**: la fonte certa ed affidabile

specifiche della qualità di sistema

1. **disponibilità**: accesso possibile sempre o all'occorrenza
2. **ripristinabilità**: ripristino dell'integrità fisica e logica dei dati in caso di guasto
3. **portabilità**: trasferimento tra ambienti tecnologici

relative sia a qualità inerente che di sistema

1. **accessibilità**: dati accessibili dalla tipologia di utenti cui ci si rivolge, anche se disabili
2. **comprensibilità**: comprensione possibile per gli utenti di riferimento
3. **conformità**: a norme, regolamenti o convenzioni
4. **efficienza**: prestazioni adeguate alle risorse ed alle esigenze
5. **precisione**: livello di dettaglio dell'informazione richiesta
6. **riservatezza**: accesso autorizzato e sicuro
7. **tracciabilità**: storia dei valori, dei legami e degli accessi

I database

I dati si archiviano, si organizzano e si gestiscono grazie alle basi di dati.

Una base di dati o **database** (abbreviato **db**) è una collezione organizzata di dati in formato digitale.

Le informazioni contenute in un database sono strutturate e collegate tra loro secondo un particolare modello logico (ad es. relazionale, gerarchico, reticolare o a oggetti). Gli utenti possono interfacciarsi database attraverso interrogazioni (anche dette **query**) o anche grazie a particolari applicazioni software dedicate.

Il modello di database viene progettato secondo le esigenze di una situazione reale che si vuole automatizzare (gestione magazzino, fatturazione, personale, ecc.).

Il database ha lo scopo di:

- preservare i dati nel tempo;
- favorirne il reperimento;
- facilitarne la condivisione.

L'elemento di base di un database è la **tabella**, ogni riga corrisponde a un **record**, ovvero un elemento del database che viene descritto mediante **attributi** che servono per una migliore comprensione dei dati, e quindi per poter assegnare loro dei significati in modo chiaro e anche leggibile da programmi (categorie, etichette, dati geografici, temporali, ecc.).

Ipotezziamo di voler organizzare un evento, avremo bisogno prima di tutto di raccogliere i dati relativi a tutte le organizzazioni che vogliamo coinvolgere.

La tabella "**organizzazioni da coinvolgere**" avrà nelle colonne gli attributi che ci interessano (denominazione, tipologia, e-mail, indirizzo, città, regione, ecc...) e nelle righe i dati delle organizzazioni che compongono la nostra mailing list.

I database relazionali

Un database relazionale è uno dei più diffusi modelli logici di rappresentazione e strutturazione dei dati ed è strutturato intorno al concetto matematico di relazione tra tabelle.

Proseguendo con l'esempio precedente, se ad esempio possiedo due tabelle "**Organizzazione da coinvolgere**" e "**Contatti**" (persone con indirizzi e recapiti, settore, ecc.) , prevedendo una colonna (un attributo) in cui metto in relazione la riga di una persona

con la riga della tabella della sua organizzazione, posso interrogare il db ed estrarre i dati filtrando i campi che mi interessano, 'pescati' quindi da due tabelle differenti.

Ad es. se voglio invitare ad un evento solo le persone degli enti locali della Puglia che si occupano di turismo, posso estrarre i dati di queste persone interrogando il database secondo i filtri: regione puglia (dalla tabella organizzazioni) + settore turismo (dalla tabella Contatti).

Quali dati

Per un'amministrazione sono tanti i dati che possono essere fondamentali per definire politiche generali, obiettivi, piani di azione, risorse.

Sempre di più è necessario **analizzare i dati in modo integrato**, considerando i fenomeni e i processi che definiscono il contesto in cui si opera.

Un esempio dalla vita professionale potrebbe essere rappresentato da un impiegato dei servizi turistici che vuole migliorare i flussi in un determinato quartiere. Per comprendere i tipi di interventi necessari deve utilizzare dati da diverse fonti:

- ambiente
- servizi sanitari
- trasporti pubblici (statici e dinamici)
- mobilità (traffico, percorrenza, parcheggi)
- catasto e mercato immobiliare
- territorio e attività economiche
- criminalità
- ricettività alberghiera
- lavoro

Tutti questi dati dovranno naturalmente essere **geolocalizzati** per realizzare un'analisi integrata.

I Geodati

Sempre più i **dati sono significativi se letti rispetto alla loro geolocalizzazione**

(collocazione geografica) ed è proprio il territorio il luogo che permette di dare significato profondo a dati di diversa fonte (es. sanitari, reddituali, ambientali).

I **geodati** sono il risultato dell'assegnazione ad una **informazione** in formato digitale dei dati relativi alla sua **localizzazione geografica** sulla superficie terrestre attraverso delle coordinate relative a un sistema di riferimento geografico.

I dati territoriali vengono definiti dal Codice dell'Amministrazione digitale come qualunque informazione geograficamente localizzata

L'Italia ha definito anche un Repertorio Nazionale dei dati territoriali che raccoglie tutti i geodati nazionali secondo le linee dettate dalla direttiva europea INSPIRE.

Le politiche sui dati per la PA

Le politiche digitali della PA stanno affermando con forza alcuni principi sulla gestione dei dati che possiamo così riassumere:

- **i dati devono essere intesi come patrimonio comune** e devono avere le seguenti caratteristiche:
 - **non replicabilità:** i dati non devono essere replicati da più amministrazioni e men che meno all'interno della stessa amministrazione.
 - **fruibilità:** i dati, devono essere fruibili da più amministrazioni, favorendone il riutilizzo con la loro esposizione in formato aperto;
 - **unicità del dato:** è garanzia della possibilità di realizzare efficacemente l'obiettivo del punto unico di accesso digitale ai servizi pubblici;
- **dati al centro dei sistemi** - il dato è **centrale** rispetto ai sistemi informativi e le applicazioni utilizzano ed elaborano dati e informazioni accedendovi direttamente;
- **open data by default** - tutti i dati per cui si prevede pubblicazione obbligatoria devono essere forniti in formato aperto. Questo è uno **stimolo al riutilizzo** da parte dei privati, cittadini e imprese, oltre che un'opportunità per ricevere feedback sulla qualità dei dati pubblicati;
- **qualità dei dati** - indispensabile sempre, perché permette di comprendere la **qualità dei processi** che li producono, e delle **basi di dati di interesse nazionale che sono considerate "critiche"**, importanti perché utilizzate da più amministrazioni sul territorio nazionale e di elevato impatto socio-economico;

- **controllo di gestione** - la costituzione di basi di dati che raccolgano tutti i principali dati sui processi delle amministrazioni è preconditione per l'attuazione di un **efficace controllo di gestione**, elemento fondamentale per l'attuazione di un costante monitoraggio della spesa e di una non estemporanea (ma sistematica) spending review.

Protezione e sicurezza

Per utilizzare la grande mole di dati che offre la rete è fondamentale l'affidabilità, perché è questa che ci permette di personalizzare i servizi sulla base delle scelte dell'utente.

Quanto più si afferma questo flusso, tanto più diventa cruciale la gestione della sicurezza sull'accesso ai dati e la loro conservazione.

In questo contesto la **protezione dei dati sensibili dei cittadini**, in relazione al loro stato, alla loro salute, diventa uno degli elementi fondanti:

- della fiducia nel digitale e nella rete;
- della sicurezza socio-economica di uno Stato.

Molti governi stanno lavorando in tal senso per adeguare le leggi in modo che considerino pienamente questo aspetto, ma naturalmente il tema deve essere affrontato anche a livello tecnico, con **programmi adeguati e internazionali per la sicurezza in rete**.

L'equilibrio da trovare, e che diventa un obiettivo da perseguire per le infrastrutture pubbliche di servizi digitali, è tra:

- il **diritto alla protezione individuale dei dati**, che spinge verso l'adozione di tecniche di crittografia sempre più spinte, anche sui social, per rendere le comunicazioni private non leggibili da terzi;
- la **necessità di evitare che le comunicazioni private di organizzazioni criminali e terroristiche** siano preservate a tal punto da sfuggire a possibili indagini.

Big Data

Per orientare correttamente lo sviluppo e la gestione dei servizi è sempre più necessario utilizzare in tempo reale dati dalle fonti più svariate.

Se ci occupiamo di una località turistica con problemi di gestione dinamica dei flussi e dell'accoglienza, possiamo essere in grado di prevedere i flussi turistici in entrata e in uscita, e modellare di conseguenza il sistema di mobilità e di accoglienza, incrociando i dati di:

- SIM telefoniche con quelli dei GPS,
- check-in
- contenuti dei social network
- acquisti dei trasporti (treni, aerei, autobus, navi)
- servizi
- attività
- ricettività

Queste analisi dei dati oggi sono possibili grazie alle applicazioni di **Data Analytics** che consentono di raccogliere ingenti volumi di dati (**Big Data**) prodotti da più applicazioni e più dispositivi e analizzarli in tempo reale.

L'obiettivo dell'elaborazione dei Big Data consiste nell'estrarre informazioni utili nell'analisi di specifici fenomeni grazie all'eterogeneità delle fonti analizzate:

- **dati destrutturati**
- **dati di formato eterogeneo** (foto, testi, video, etc.)
- **dati di varia provenienza** (Internet, social network, SIM telefoniche, GPS, transazioni, e-commerce, etc.).

La rilevazione statistica non riesce a fornire risultati soddisfacenti nel caso di analisi di dati molto corposi, dinamici, destrutturati o relativi a tendenze, percezioni, o su diversi range temporali.

Con l'analisi dei Big Data invece è possibile:

- indagare oggi fenomeni passati, per confrontarli con eventi attuali;
- effettuare analisi in modo veloce, immediato, flessibile, con dati aggiornati in tempo reale;
- identificare segnali deboli (l'indagine si modella rispetto ai dati che fluiscono);
- effettuare previsioni accurate di eventi futuri.

Questo tipo di analisi è necessaria per una gestione dinamica del lavoro orientata all'innovazione, con feedback rapidi.

Le nuove figure professionali correlate

Data la complessità e la dinamicità dei fenomeni basati sui dati, a un e-leader non si richiede soltanto di comprendere come utilizzare la ricchezza dei dati per la definizione di obiettivi, strategie, gestione del miglioramento e il monitoraggio della propria organizzazione, ma anche di dotarsi di figure professionali con competenze specifiche sul tema.

Questo significa favorire lo sviluppo di competenze per la formazione di figure che possano dare un contributo adeguato per la valorizzazione dei dati, come ad esempio:

- **Data scientist**, in grado di realizzare elaborazioni e analisi accurate sui Big Data;
- **Geographic Information Specialist**, in grado di organizzare in modo organico l'informazione geografica;
- **Geographic Information Manager**, in grado di far sì che tutte le strategie e le azioni di un'amministrazione possano trarre pieno beneficio dall'utilizzo dei geodati;
- **Chief Data Officer**, in grado di indirizzare e gestire l'architettura dei dati in un'organizzazione, centro nevralgico del suo funzionamento.

Nel 2015 solo il 17% delle imprese italiane si è dotata di un Chief Data Officer e solo il 13% di un Data Scientist.

La misurazione per il miglioramento

I dati sono necessari anche per definire **obiettivi e azioni**, e per migliorare le **performance**.

Per migliorare le performance dei processi organizzativi, è indispensabile che alle attività di **pianificazione e controllo** seguano delle **attività di valutazione** che confrontino il comportamento e i risultati osservati con quelli attesi, così da poter intervenire con eventuali modifiche nella definizione, e nella scelta degli strumenti e delle risorse necessarie.

Il confronto è però possibile solo se si dispongono di **dati affidabili** della situazione di partenza (che permette di definire in modo accurato e realistico gli obiettivi) e di quella osservata.

Poiché **si può migliorare solo ciò che si misura, senza dati non è possibile definire attività di miglioramento**.