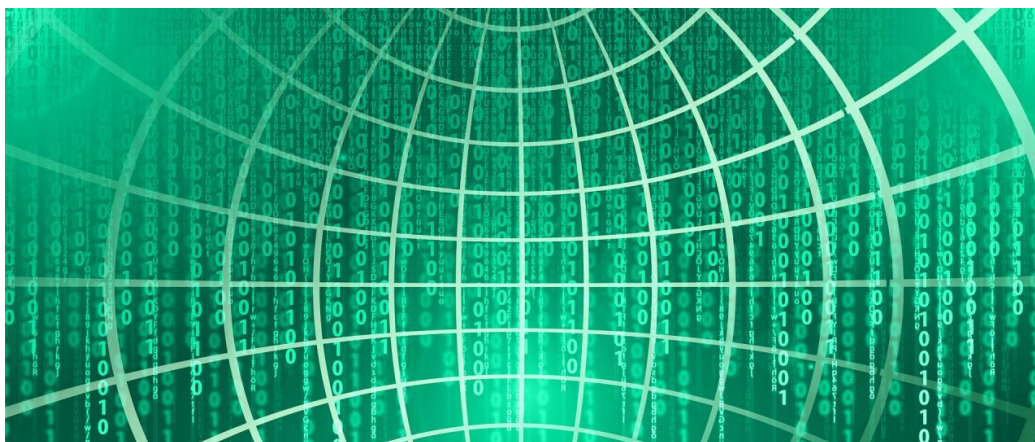




Link: <http://www.datamanager.it/2021/04/data-quality-universo-a-otto-dimensioni/>



13 aprile 2021 - Di Giuseppe Mariggìo

**ESTRATTO**

## **Data quality, universo a otto dimensioni**

***Completezza, accuratezza, tempestività, coerenza, univocità, integrità, conformità, ma anche tracciabilità dei processi di trasformazione dei dati. Come coltivare la qualità dei dati per evitare il rischio di decisioni sbagliate, processi inefficienti, performance ridotte, opportunità mancate e sfiducia***

La qualità dei dati è un requisito fondamentale per ottimizzare la raccolta e l'analisi dei dati. Oltre che la successiva esplorazione e interpretazione delle informazioni, nascoste nei Big Data. Un corretto processo di data quality si traduce in aumento dell'efficacia dei processi decisionali. I dati "puliti" sono la base per sviluppare progetti efficienti e solidi, in grado di supportare appieno le strategie data-driven. Si fa presto a dire analytics, monetization e trasformazione dei dati in conoscenza. Occorre però essere meno superficiali nell'associare ai dati un valore intrinseco. Nelle operazioni di utilizzo, secondo i diversi paradigmi, i dati vengono conservati, spostati o copiati da un repository all'altro. Tutte operazioni – costose in termini di risorse – che rischiano di avere un impatto negativo dal punto di vista dell'integrità e della qualità dei dati.

Senza una base integra, univoca e accessibile di dati, è difficile estrarre vero valore di business. Il ciclo di vita del dato, dalla gestione all'analisi, richiede pratiche e strumenti adeguati. Sicurezza, consistenza, incorruttibilità, qualità e integrità dei dati sono fattori determinanti per le strategie di ogni organizzazione, soprattutto a supporto dei processi decisionali assistiti dall'intelligenza artificiale. Non basta avere una fonte di informazione, è necessario porre al centro delle attività aziendali una vera "data community" per incentivare un'evoluzione più spedita e stabile interna all'organizzazione. Troppo spesso le aziende si ritrovano a lavorare con dati non "puliti". Questa non uniformità del dato può avere impatti potenzialmente disastrosi per chi necessita di muoversi con agilità e velocità nel proprio mercato di riferimento. Oggi, i CDO sono chiamati a colmare il divario fra l'IT e il mondo del business, non solo attraverso l'inserimento di soluzioni innovative o di risorse professionali specializzate, ma soprattutto con la possibilità di condividere scelte, esperienze e responsabilità.

## TRACCIABILITÀ DEI PROCESSI

La qualità del dato si misura in base a sette dimensioni: completezza, accuratezza, tempestività, coerenza, univocità, integrità, conformità. Ognuna di esse deve essere rispettata per avere un dato di qualità. Oggi, il concetto di data quality sta evolvendo verso una ottava dimensione che è essenzialmente quella legata alla tracciabilità dei processi di trasformazione dei dati, come conseguenza dei nuovi algoritmi. «Il data lineage è indispensabile per tenere traccia del confine, talvolta sottile e ingannevole, tra dato, informazione e insight» – spiega **Giancarlo Vercellino, associate director Research & Consulting di IDC Italy**. «Un elemento indispensabile per fare quel “sanity check” che è richiesto per valutare la correttezza dei workflow analitici». I dati sono alla base, ma le aziende come possono fidarsi completamente dei dati in loro possesso? Secondo Vercellino – le aziende non devono fidarsi dei dati, ma dei processi che sono in grado di costruire sui dati. «I processi possono essere monitorati e se necessario emendati. I dati, soprattutto quando provengono da fonti esterne, non possono sempre e comunque essere validati o accertati. Una azienda propriamente data-driven deve imparare a convivere con l'errore, che è un po' il rumore di fondo, il rumore bianco, che contraddistingue una società che produce e macina grandi quantità di dati ogni giorno».

L'automazione di tanti processi che caratterizzano il back-end dell'IT aziendale – oggi parliamo di data governance, ma pensiamo a tutto quanto sta avvenendo con il Serverless lato configurazione e deployment – avrà un impatto molto importante nel rafforzare la credibilità di una organizzazione data-driven. «L'errore umano – spiega Vercellino – è il primo fattore che può mettere a repentaglio la qualità dei dati, introducendo distorsioni e manipolazioni che possono poi impattare sulle decisioni finali. Una macchina capace di portare avanti in modo autonomo e controllabile alcuni aspetti legati alla data curation potrebbe accelerare in modo importante il time-to-insight».

Una comprensione approssimativa della realtà porta sempre a decisioni errate. Quando gli insights vengono utilizzati per guidare i manager esecutivi e i direttori di funzione, si entra nel dominio della politica aziendale e capire se una decisione sia giusta o sbagliata diventa un tema di dibattito o un esercizio di potere. «Quando gli insights vengono impiegati per automatizzare i processi – precisa Vercellino – se sono stati commessi degli errori di gestione del dato, diventa immediatamente evidente a tutti: in primo luogo ai clienti, poi a chi lavora a contatto con i clienti e quindi a chi controlla i numeri. In questo caso, diventa davvero difficile nascondere la polvere sotto il tappeto». L'implementazione di un attento processo di data quality può ridurre costi e rischi su entrambi i fronti. Le ragioni sono le più disparate e dipendono dai casi d'uso. «In ambito finanziario – spiega Vercellino – la qualità dei dati può mitigare i rischi di un affidamento. In ambito di servizio, la qualità dei dati può ridurre il rischio di churn dei clienti. In ambito industriale, la qualità del dato può garantire l'efficienza del prezzo edonico di un prodotto, calcolando finemente quello che è il contributo marginale di ogni fattore produttivo alla qualità del prodotto finale. I casi possibili sono tanti». Il CDO ha la responsabilità di costruire un processo capace di gestire gli analytics e gli insights come se fossero un processo industriale – avverte Vercellino – garantendo la valorizzazione dei dati e promuovendo una cultura centrata sul dato a tutti i livelli aziendali. «Si tratta di compiti assolutamente non facili che richiedono una stretta collaborazione con CEO, CIO, CTO, COO e CMO per essere perfettamente allineati sugli obiettivi e sui tempi necessari per una trasformazione di questo tipo».

...

## EFFICACIA E COERENZA

Per **Mirko Menecali, Partner and Alliance manager di Sinfo One** i dati sono il patrimonio principale di qualsiasi sistema e la loro qualità è direttamente connessa con l'efficacia complessiva del sistema/processo, sia esso un'applicazione operativa, un ERP, un tool di analitica e business intelligence. E oltre ai requisiti di completezza, precisione, disponibilità e tempestività, ce ne sono molti altri da prendere in considerazione – *«come la coerenza centrale nei sistemi di analitica che usano diverse fonti dati, l'univocità nei sistemi ERP, la conformità formale fondamentale nei sistemi EDI, l'integrità fondamentale quando si usa IoT»*.

Difficile dire quale sia il più importante, anche se negli ultimi tempi viene stressato il concetto di tempestività intesa come velocità di fruizione. Entrando nello specifico delle applicazioni di BI – Menecali spiega che gli utenti evoluti, come i data scientist, devono accedere a importanti moli di dati preparati per supportare analisi complesse e destrutturate. *«Rispondere a queste richieste in modo tempestivo richiede, anche in aree tradizionali come il data warehousing, approcci evoluti come l'Unified Star Schema e l'utilizzo di "tabelle ponte" per aumentare la tempestività di presentazione dei dati»*.

...

## DATA QUALITY AS A SERVICE

Ogni progetto di implementazione deve includere un'adeguata fase di data quality – afferma **Mirko Menecali di Sinfo One**. *«Noi adottiamo una metodologia che impone determinati standard aiutandoci a non sottovalutare questo processo. Alcuni clienti hanno adottato questi standard procedurali, per esempio, individuando figure specifiche a livello di funzione responsabilizzate sulla qualità dei dati generati e figure centrali con il compito di gestire il governo delle informazioni»*.

...