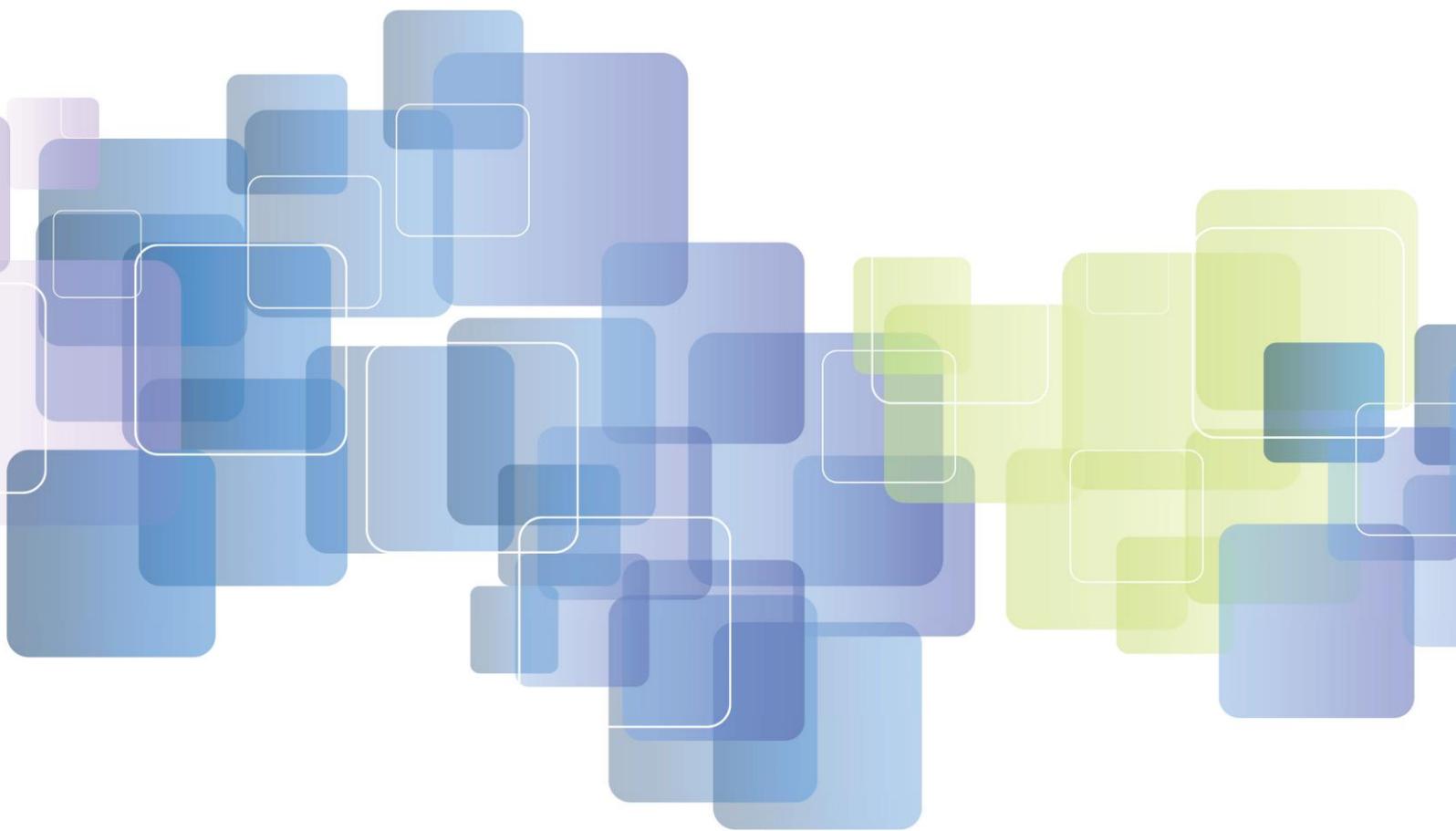




Data Warehouse: una sola lingua

Iolanda Salinari



Il Glossario: uno strumento prezioso

Uno dei fattori critici per la realizzazione del data warehouse è costituito dalle difficoltà che si incontrano per integrare i dati aziendali, partendo dai sistemi legacy. Il glossario dei termini più ricorrenti del business può giocare un ruolo fondamentale per stabilire un vocabolario e una base di conoscenza comuni tra aree funzionali diverse. In un'ottica più generale di condivisione della conoscenza in azienda, può rappresentare uno strumento prezioso per qualsiasi tipo di progetto.

La definizione dei dati

Il linguaggio naturale è spesso impreciso, sovente una parola ha significati diversi e lo stesso significato può essere espresso con parole differenti. Sono stati scritti centinaia di libri che trattano delle problematiche relative alla comunicazione con il linguaggio naturale; è scontato che questi problemi si riflettano pesantemente anche nei nostri sistemi informativi: quante volte, in una stessa applicazione, un dato è utilizzato per significare differenti cose oppure più dati sono impiegati per indicare lo stesso concetto, lo stesso oggetto del mondo reale?

Il linguaggio e i dialetti aziendali

Se estendiamo poi i confini della nostra analisi dalla singola applicazione all'intera azienda, i problemi si ampliano e si moltiplicano; l'organizzazione procedurale degli attuali sistemi informativi, che, quasi sempre, rispecchia rigidamente quella delle funzioni aziendali, ha consolidato nel tempo vocabolari (terminologie) propri delle diverse aree funzionali, talvolta in conflitto tra loro. Spesso emerge la necessità di fare chiarezza sul significato di informazioni fondamentali quali i clienti, i prodotti, le entità geografiche, proprio quelle informazioni che necessitano di maggior condivisione in azienda. In molti casi l'assenza di chiarezza è conseguenza di un utilizzo improprio delle informazioni da parte dei sistemi applicativi per far fronte a problemi contingenti (nuovi requisiti, modifiche urgenti), il significato originario dei dati si è via via modificato con il passare del tempo. Così, ad esempio, alcuni sistemi informativi dell'area Vendite, trattano allo stesso modo unità organizzative interne e clienti esterni, con l'obiettivo di semplificare i processi elaborativi, ma in un data warehouse la distinzione diventa indispensabile perché gli utenti possano differenziare opportunamente tra clienti interni e clienti esterni. Più in generale, nell'ottica di un'organizzazione sistemica dell'azienda, diventa necessario che gli oggetti di business e i loro attributi siano chiaramente identificati e definiti in modo adeguato, per consentire all'utente un utilizzo efficiente ed efficace delle informazioni aziendali.

L'importanza di una corretta definizione dei dati nel data warehouse

L'impatto di una definizione inadeguata è molto più grande in un data warehouse che in un sistema operativo. In un sistema operativo è sufficiente che il progettista capisca il significato dei dati per utilizzarli correttamente nei programmi.

Gli utenti non sono direttamente interessati alla definizione di ogni oggetto e di ogni dato elementare, è sufficiente che questi siano opportunamente "etichettati" nei report e nelle videate.

In un data warehouse la situazione è ben diversa per almeno due motivi:

1. l'utente finale può trattare direttamente i dati ed è quindi molto importante che capisca il vero significato di ogni dato elementare
2. sovente i dati trattati provengono da sistemi alimentanti diversi, appartenenti a differenti aree funzionali e la loro integrazione non è cosa semplice.

La parola d'ordine è integrazione

L'integrazione diventa un prerequisito fondamentale del data warehouse: all'interno del sistema deve essere presente una definizione univoca di un certo dato, anche a fronte di diversi utilizzi.

A questo scopo, nella fase di analisi del data warehouse si dovrà chiarire il significato del dato e gli eventuali meccanismi di conversione fra le varie rappresentazioni dello stesso presenti nei diversi sistemi alimentanti: l'analisi dei dati dei sistemi sorgente è uno step cruciale nella costruzione del data warehouse. In questa fase occorre risolvere tutti i problemi legati a campi o strutture dati che possono assumere significati diversi a seconda del contesto o a campi con nomi fuorvianti.

Le definizioni di business degli elementi del data warehouse (metadati) sono create come parte dello sforzo di modellazione: nel modello logico definiremo entità e associazioni tra entità sulla base delle regole di business, per ogni entità identificheremo gli attributi, determinando il significato di ognuno di essi. Gli utenti finali otterranno le definizioni di dimensioni e fatti del data warehouse attraverso i metadati.

I metadati: un problema apparentemente irrisolvibile

Ovviamente lo sforzo di modellazione è tanto più oneroso quanto più è carente la documentazione dei sistemi di base. Così si paga lo scotto di un mancato impiego della modellazione a tutti i livelli in azienda; purtroppo anche nei casi migliori, laddove esiste un modello dei dati /processi per una certa applicazione, spesso non è prevista alcuna integrazione in un modello aziendale, che consenta di analizzare la situazione attuale secondo una visione di business.

Per dirla con Inmon, i metadati sono stati per molto tempo "the Wednesday's child" dei sistemi informativi. Dopo i fallimenti dei primi dizionari dati degli anni '70 fino alla leggendaria disfatta del repository IBM all'inizio degli anni '90, i metadati hanno rappresentato per la comunità informatica un problema apparentemente irrisolvibile. Se da un lato, dalla metà degli anni '80, c'è stato un notevole sforzo in molte aziende per uniformare il linguaggio, integrare e rendere consistenti i metadati all'interno dell'organizzazione, dall'altro lato i "nuovi" potenti tools messi a disposizione di progettisti (ad es. ambienti per la programmazione visuale) e di utenti finali (spreadsheet, . . .), hanno incoraggiato un ambiente di creatività e spontaneità, in assoluto contrasto con la disciplina di un approccio centralizzato alla gestione dei metadati.

D'altra parte, in passato, è sempre stato arduo dimostrare al business i benefici di un repository centrale, di una "mappa" delle informazioni aziendali. La Data Administration o funzioni affini, nate per gestire centralmente la conoscenza dei dati aziendali, anche se molto pubblicizzate, non hanno trovato applicazione su larga scala.

Le aziende non ne comprendevano bene il reale significato, anche perché a proporle è stata quasi sempre l'area tecnologica dei sistemi informativi, che ne ha dato troppo spesso connotazioni molto legate alla tecnologia e non ha saputo chiarire all'alta direzione i suoi veri obiettivi e i benefici risultanti per il business aziendale.

I metadati: che cosa è cambiato oggi?

Oggi, diventa sempre più evidente che per sfruttare al meglio le opportunità di business, è necessario disporre velocemente di informazioni accurate, aggiornate e significative, di conseguenza il ruolo dei metadati nei sistemi di analisi del business e di supporto alle decisioni aziendali diventa fondamentale: solo un'adeguata conoscenza delle informazioni aziendali ne permette un utilizzo efficiente da parte degli utenti. Ed è appunto dai metadati che l'utente finale ottiene le definizioni dei dati. Nella forma più semplice, i metadati possono essere descritti in un manuale, che contiene le definizioni e altre informazioni correlate ai dati. Un'altra possibilità consiste nel fornire un accesso elettronico ai metadati, con l'utilizzo di strumenti (servizi) più o meno complessi e articolati, quali query, directory dei metadati e help. Indipendentemente dagli strumenti utilizzati, i dati elementari nel data warehouse devono essere definiti in modo chiaro, e gli utenti finali devono poter accedere facilmente a tali definizioni per essere messi in grado di capire il contenuto del data warehouse e usarlo in maniera efficace per prendere le decisioni strategiche nel modo migliore.

Il glossario: un primo livello di metamodello

Nelle fasi precoci del progetto, quando si raccolgono le esigenze degli utenti, si identificano i sistemi alimentanti e i dati necessari, si definiscono in bozza i modelli dati, utenti e processi di alto livello, un glossario, dove raccogliere i termini principali dell'area di analisi, può assumere un ruolo basilare per stabilire un vocabolario comune tra utenti e gruppo di progetto e rappresentare un primo passo per la costruzione del metamodello, per assicurare il consenso sulle definizioni dei metadati e sulla consistenza del modello del data warehouse. Il glossario raccoglierà in prima istanza i termini di business relativi a dimensioni e fatti contenuti nel data warehouse, ogni termine sarà accompagnato dalla dichiarazione del significato ed eventualmente da altre osservazioni, quali regole di business, regole di calcolo, fonte e proprietari dell'informazione, ecc... Volendo, potremo poi aggiungere a questa raccolta le definizioni di altri oggetti di business (unità organizzative, ecc..), e tutto quanto contribuisce a fornire un vocabolario, una base di conoscenza comune e condivisa. Articolando il glossario in diverse sezioni, potremo unire ad esempio una sezione relativa agli acronimi maggiormente utilizzati nel dominio di analisi del progetto ed eventualmente una per i termini tecnici più ricorrenti in ambito data warehouse.

Un esempio di glossario

Forniamo un [esempio di glossario](#), costruito seguendo il modello dell'autorevole Glossario di Unified Modeling Language di Booch, Rumabaugh e Jacobson. In esso, la definizione dei diversi elementi avviene ad un livello di astrazione più alto rispetto all'individuazione tecnica e relativa descrizione di entità e attributi, consentendo tuttavia di chiarire il significato di business delle informazioni indipendentemente dalle diverse rappresentazioni fisiche. Se il dominio di analisi copre più aree funzionali (Marketing, Produzione, Amministrazione, . . .), il glossario diventa anche l'occasione per integrare terminologie diverse, individuare sinonimie e divergenze, chiarire il significato delle informazioni al di là dei differenti utilizzi con gli attori rappresentanti delle varie aree funzionali. In assenza di un modello dei dati aziendali, il glossario può essere uno strumento di banale semplicità, ma molto efficace per "mappare" le conoscenze via via raccolte nei progetti in essere, per evidenziare gli oggetti ricorrenti nel business e le informazioni più importanti, focalizzandosi fin dall'inizio sugli elementi utili per una futura integrazione a livello aziendale. Se, invece, già esiste una "mappa" dei dati aziendali, il glossario può fornire un'utile vista di sintesi, limitata alle informazioni di maggior interesse.

Una volta scritto, pubblicarlo!

Ovviamente, è importante che la conoscenza acquisita, spesso con non poche difficoltà per integrare informazioni provenienti da realtà diverse, non resti confinata al progetto e ai suoi partecipanti, ma sia condivisa e diffusa in azienda. A questo scopo niente si presta meglio della tecnologia Internet quale infrastruttura di comunicazione, per la sua semplicità d'uso, l'agevole struttura di navigazione e la flessibilità con cui possono essere organizzate e presentate le informazioni. Si tratta di pubblicare il nostro glossario su pagine Web, in modo da sfruttare i motori di ricerca e la possibilità di legami ipertestuali all'interno del documento (si può "puntare", ad esempio, da un termine ad un suo sinonimo) o con altre risorse informative e di renderlo opportunamente disponibile tramite l'intranet aziendale. A questo punto potremo considerarlo una buona base di partenza per la definizione di un glossario aziendale.

E' possibile fare il download del template per il [glossario dei termini](#)



Tecnet Dati s.r.l.
C.so Svizzera 185 -
10149 - Torino (TO), Italia
Tel.: +39 011 7718090 Fax.: +39 011 7718092
P.I. 05793500017 C.F. 09205650154
www.tecnetdati.com

