



# Data Governance

## Approccio al progetto e componenti tecnologiche

**F. Vettore**

*IBM SWG Industry Solutions*

**E. Caraffi**

*IBM SWG IM Tech Sales*

IBM Information  
On Demand 2008  
>>> Comes To You

*ALLA LUCE DELL'INFORMATION ON DEMAND*

*Milano, 15 aprile 2008*



# Le informazioni sono l'asset portante delle organizzazioni



- Insieme a Human e Financial Capital, le Informazioni sono ormai un vero e proprio patrimonio.
- La gestione di questo patrimonio informativo presenta notevoli margini di miglioramento: gran parte dei problemi riguardano aspetti di Conoscenza e Qualità.

**60% + of CEOs:** Need to do a better job capturing and understanding information rapidly in order to make swift business decisions

Only **1/3** of CFOs believe that the information is easy to use, tailored, cost effective or integrated

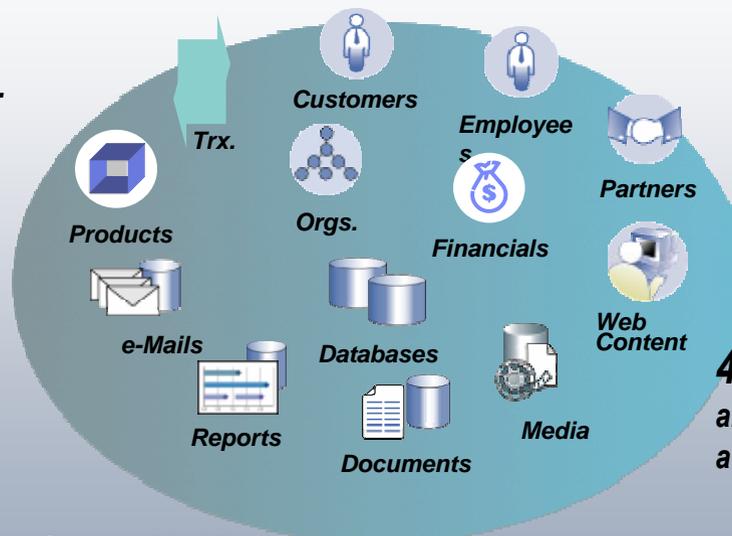
**79%** of companies: have 2 + repositories...  
25%: have 15 +

**30%** of people's time is spent searching for relevant information

**85%** of information is unstructured

**17%** of IT Budgets Spent on Storage HW, SW, People

**37%** CGR Disk Storage Growth '96-'07



**30-50%** of design time is copy management

**48** disparate financial sys. and 2.7 ERP sys. in the average \$1B company

**40%** of IT budgets spent on data integration

Sources: IBM Attributes & Capabilities Study, 2005; Client Interviews 2004, IBM CFO Survey, 2006

# I principali problemi legati alla Conoscenza



- Documentazione **incompleta e non aggiornata, settoriale, frammentata e volatile** del patrimonio informativo.
- Conoscenze **concentrate su poche persone**.
- Manutenzione delle informazioni **pesante e “time consuming”**.
- Ridondanza e proliferazione delle stesse informazioni: **basso riuso e poca standardizzazione**.
- **Definizioni semantiche imprecise**, non rigorose e ambigue, mancanza di standard.
- Sicurezza: distribuzione e accesso **non sufficientemente controllati**.

# I principali problemi legati alla Qualità dei dati



- I dati non sono **accurati e precisi**.
- Dati **non coerenti e non intelligibili**.
- Informazioni disponibili spesso **solo a livello aggregato**.
- Informazioni **non contestualizzate nel tempo**.
- Gestione non corretta della **storicità dei dati**.
- Decisioni spesso basate su dati e metadati **vecchi o forniti in ritardo**.
- Progetti IT inquinati in partenza da **dati inesatti**.

# I principali problemi legati a Conoscenza e Qualità dei dati



- Tali problemi si sono aggravati negli ultimi anni a causa di:
  - **consolidamenti,**
  - **crescita di complessità/eterogeneità,**
  - **compliance e crescita dei volumi**
  
- Le aziende ammettono di :
  - non essere certe della propria capacità di aderire alle normative
  - non essere in grado di far leva in modo efficace sul patrimonio informativo dei propri clienti
  - di aver costruito delle soluzioni verticali per i sistemi di gestione
  
- Il costo della **scarsa qualità nei dati** è stato stimato, a più riprese e da più fonti, essere compreso fra il **15 e il 25% del margine operativo** (Olsen, 2003)
  
- Vi è un incremento progressivo continuo delle richieste di risorse all'IT

# I Metadati come risposta ai problemi di Conoscenza e Qualità



- I Meta Dati sono **informazioni strutturate che descrivono, esplicano, localizzano qualsiasi cosa che renda più facile recuperare, usare o gestire una risorsa informativa.**
- Spesso vengono definiti come “dati relativi ai dati” o “informazioni relative alle informazioni”.

## Meta Dati per l'IT Department (tecnici).

Le strutture IT necessitano di gestire sempre meglio i sistemi man mano che le esigenze del business crescono e diventano più sofisticate.



**Utenti tecnici**



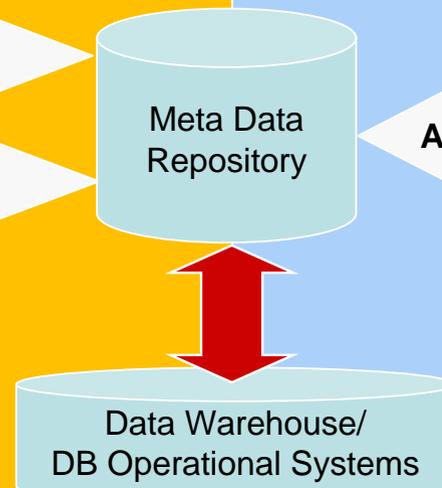
**Project Manager**

## Meta Dati per il Business.

Forniscono lo strato semantico tra i sistemi informatici (operazionali e di BI) ed i loro utenti.



**Analisti Business**



# I benefici dell'utilizzo dei Metadati



- L'importanza e la diffusione dei Meta Dati, tecnici e business, è oggi evidente in tutti gli ambiti applicativi e tecnologici.
- Il salto evolutivo che molte aziende si apprestano oggi a compiere è in direzione dell'implementazione di una struttura ufficiale di **Data Governance**.

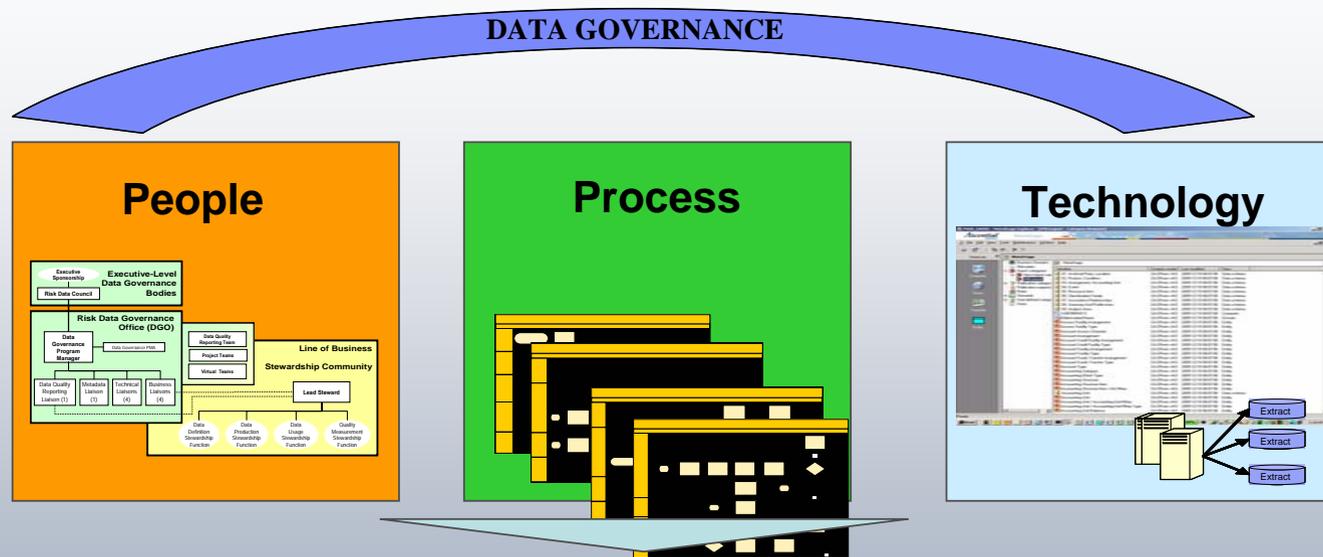
- Riducono la probabilità del fallimento di progetto
- Velocizzano il time-to-market dei sistemi
- Diminuiscono il ciclo di sviluppo dei sistemi
- Limitano la ridondanza di dati e processi
- Mettono a fattor comune il lavoro fatto dai vari team
- Diminuiscono la quota di rifacimenti
- Abbassano l'impatto del turnover di personale

- Riducono i costi per il training
- Rendono le informazioni dei DW/DM più preziose dato che migliorano la profittabilità delle decisioni degli analisti business
- Riducono le decisioni errate
- Supportano gli analisti nel trovare le giuste informazioni e nel tempo richiesto
- Aumentano la fiducia nei sistemi IT

# Data Governance



*L'espressione Data Governance si riferisce al complesso di persone, processi e tecnologie strutturati in modo adeguato ad uno sfruttamento efficace ed efficiente dei dati come risorsa e valore aziendale.*



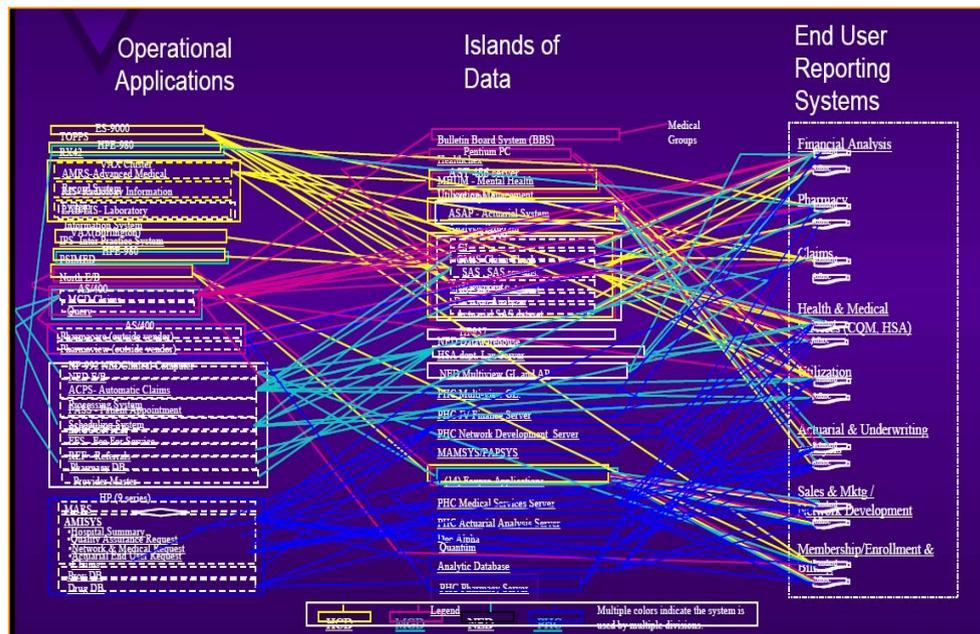
I principali obiettivi del governo dei dati sono:

- Guidare e supportare I processi decisionali
- Assicurare che le informazioni siano consistenti, diffuse e comprese
- Incrementare la fiducia nell'affidabilità dei dati e il loro uso come risorsa e valore aziendale

# Motivazione della Data Governance



Un approccio non strutturato e frammentato alla gestione dei dati inibisce la capacità delle aziende di rispondere adeguatamente ai cambiamenti.



Mancanza di flessibilità e agilità sistemi IT



Ostacoli alla capacità di gestire le informazioni



Barriere all'esecuzione delle strategie di business

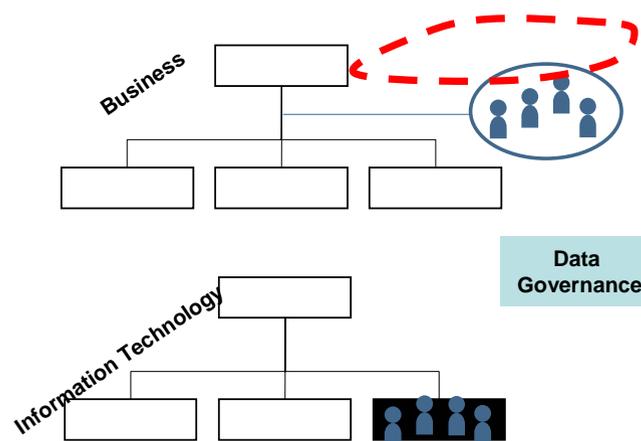
# L'approccio proposto da IBM



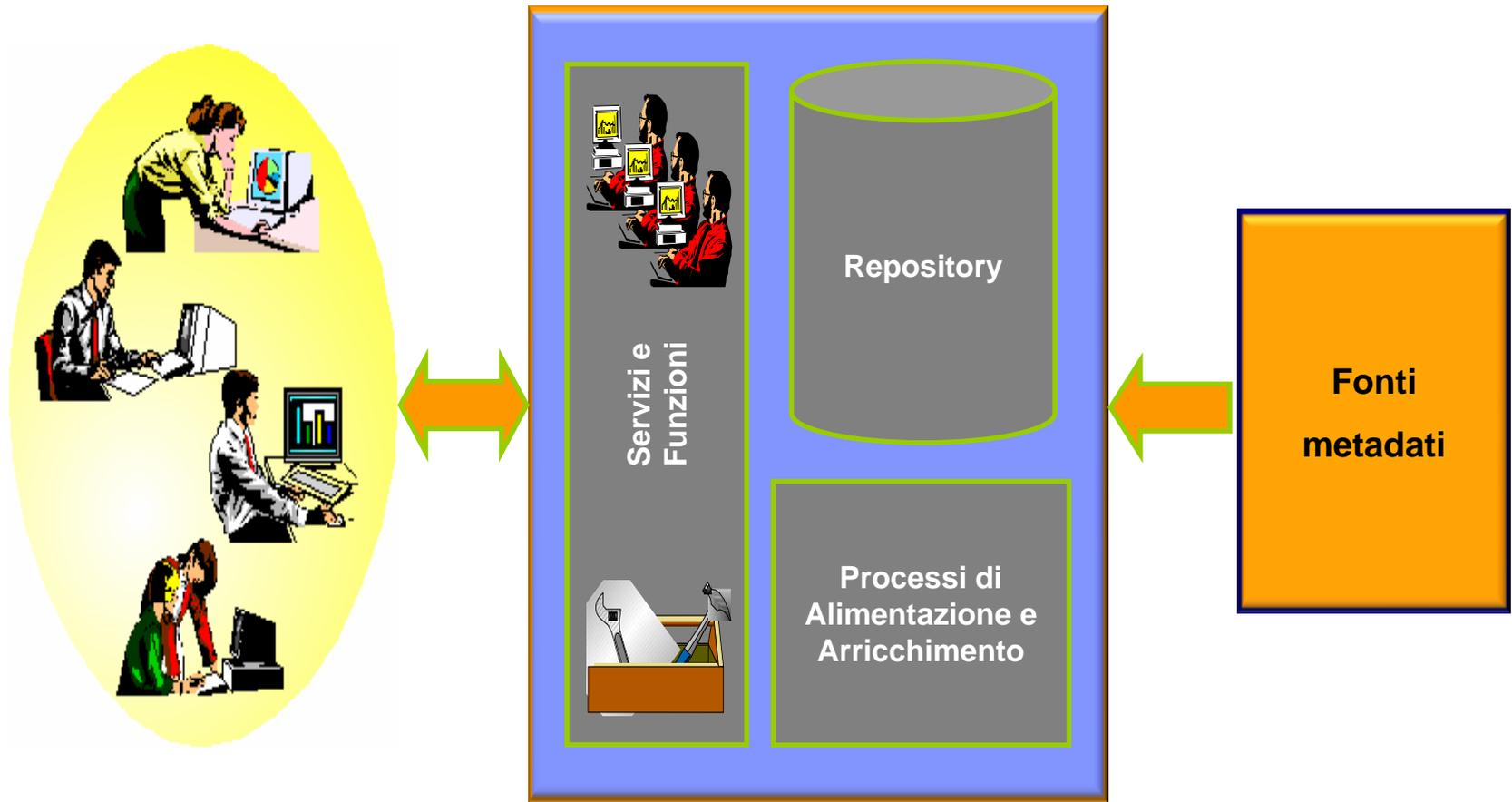
Un approccio strategico alla gestione dei dati aziendali richiede la predisposizione di una soluzione innovativa con un significativo risvolto organizzativo, supportato da strumenti, metodologie e competenze.

Modello organizzativo, composto da:

- una struttura centralizzata,
- supportata da un modello di processo che eroga servizi a tutta l'azienda,
- presidiata da un gruppo di persone con ampie competenze applicative e tecnologiche, responsabile della:
  - *condivisione della conoscenza*
  - *qualità del patrimonio informativo*



# Componenti tecnologiche della soluzione



# Servizi e Funzioni



- Repository per acquisire conoscenza su determinati elementi di business:
- Glossario su elementi di business in genere
- Elementi del modello di riferimento e di quelli banca
- Regole di trasformazione
- Contesti di utilizzo (subject area/categorie)
- Normative

- Interventi di correzione su modelli esistenti (rimozione ridondanze e ambiguità - micro) sulla base del modello di riferimento (gap analysis)
- Interventi di consolidamento (fusione di modelli di analisi)
- Razionalizzazione dei processi di elaborazione dati (economia di scala e scopo)

- Identificazione degli interventi utilizzando il modello di riferimento (gap analysis per sviluppo di nuovi progetti)
- Definizione coerente e implementazione ottimizzata di nuovi metadati di business e tecnici (modelli, regole,...)

**Knowledge management**

**Razionalizzazione dei dati**

**Costruzione**

**Filiera**

**Cleansing**

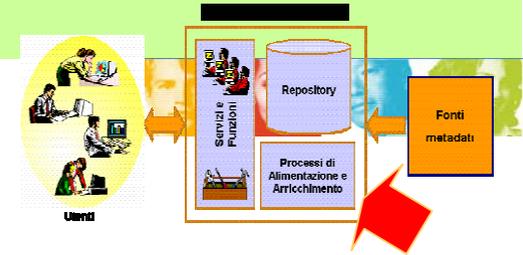
**Distribuzione**

- Analisi di impatto
- Data lineage – ricostruzione del percorso di trasformazione e derivazione dei metadati

- Intercettazione delle anomalie nella base dati e nei processi di trasformazione (diagnosi)
- Supporto interventi di bonifica della base dati
- Interventi sui processi di trasformazione per rimuovere anomalie
- Applicazione di regole di certificazione

- Creazione di viste parziali sui dati
- Profilazione degli utenti/applicazioni abilitati alle viste
- Distribuzione delle viste ai profili abilitati

# Processi di Alimentazione e Arricchimento



- Identificazione delle subject area, delle fonti : legacy, datawarehouse, datamart, presentation tool
- Acquisizione Modelli Dati (dati tecnici logici e fisici)
- Acquisizione metadati non modellati (copy cobol, declib PL1 ...)
- Acquisizione metadati non modellati da programmi di qualsiasi tipo
- Acquisizione Data Element Banca (dati tecnici e business)
- Acquisizione di dati e programmi esistenti

**Recepimento modelli e strutture esistenti (dati e programmi)**

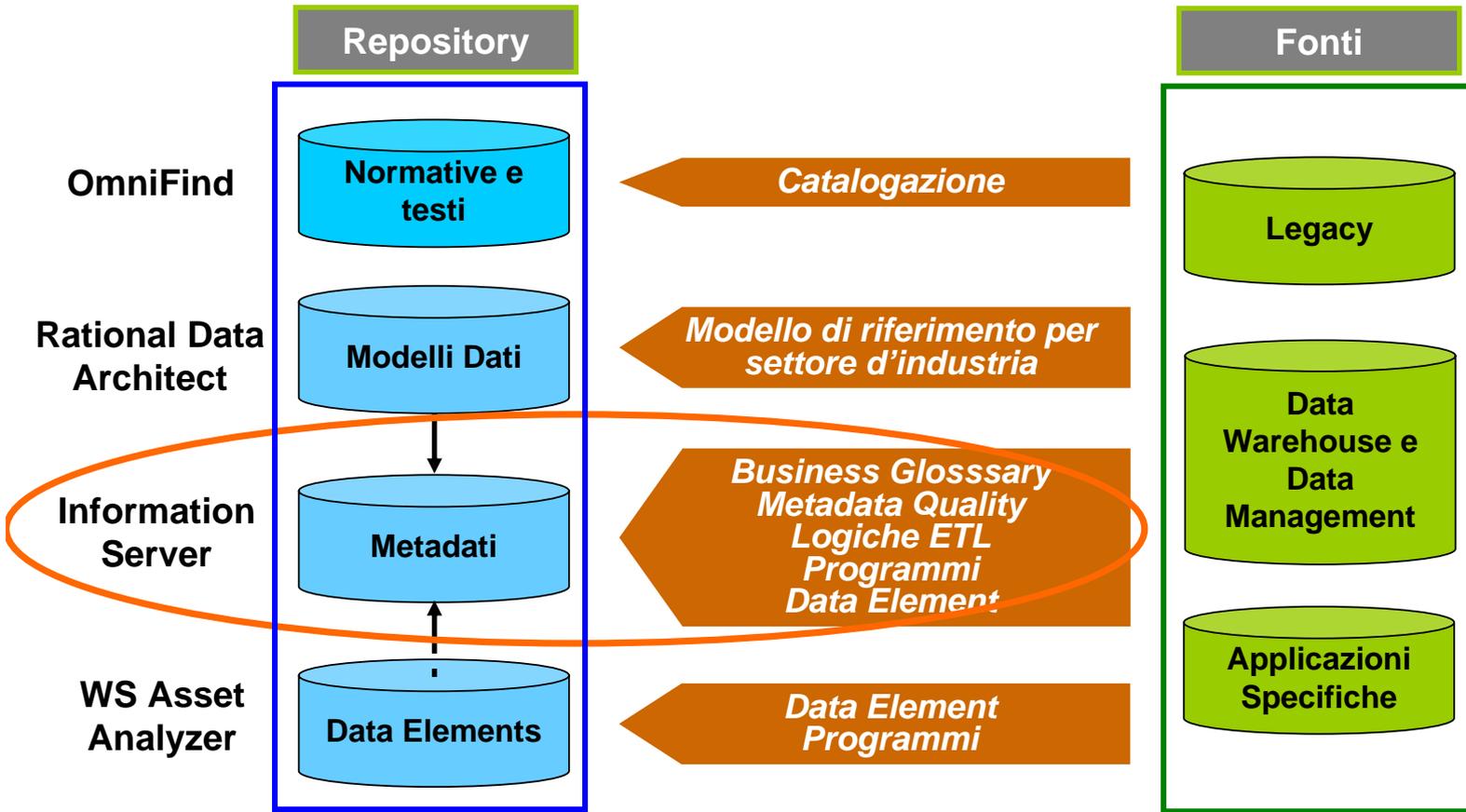
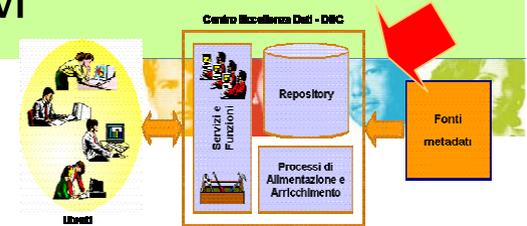
**Regole di trasformazione**

**Arricchimento del patrimonio informativo di base**

- Acquisizione regole di trasformazione tra fonti legacy e DW di primo livello (ETL e pgm ad hoc)
- Acquisizione di regole di trasf. tra DW di primo livello e DM
- Acquisizione di regole di trasformazione tra fonti e presentation tools

- Controllo/certificazione su modelli Banca e metadati (tecnici e business).
- Mapping/Compare tra modelli Banca e Modello di Riferimento
- KPI building di Data Quality della base dati mediante regole precostituite, regole del modello di riferimento o di business.
- Funzionalità di Data Quality con regole impostabili
- Arricchimento/Catalogazione del Business Glossary
- Catalogazione Data Elements da Programmi
- Catalogazione di normative interne ed esterne
- Acquisizione del Modello di Riferimento.

# Repository di Metadati: le componenti SW IBM



# I benefici della soluzione



## Conoscenza del patrimonio informativo aziendale

- Documentazione del patrimonio informativo aziendale
- Supporto alla standardizzazione e al riuso dei dati
- Supporto alla definizione dell'architettura dati
- Divulgazione della conoscenza del patrimonio informativo

## Qualità e certificazione dei dati

- Misura e valutazione della qualità dei dati
- Certificazione dei processi di trattamento dati
- Supporto al processo di quadratura dei dati

## Benefici

**EFFICIENZA**

**SEMPLIFICAZIONE**

**TEMPESTIVITA'**

**VISIONE AZIENDALE**

**TRASPARENZA**

**COERENZA E AFFIDABILITA'**

**COMPLIANCE**



# Thank YOU

IBM Information  
On Demand 2008  
>>> Comes To You

*ALLA LUCE DELL'INFORMATION ON DEMAND*

Milano, 15 aprile 2008





# IBM Information Server

## la componente Business Glossary

IBM Information  
On Demand 2008  
>>> Comes To You

ALLA LUCE DELL'INFORMATION ON DEMAND

Milano, 15 aprile 2008

*E. Caraffi*  
*IBM SWG IM Tech Sales*



# Quante culture diverse in un'organizzazione?



## Cultura di Business

## Cultura Tecnica

### Mastro Contabile

Identificativo del conto contabile anche detto "conto di contabilità"

Si compone di tre livelli :

FFFIIIVVVV.



Esperto di Contabilità

### Classificazione del Mastro Contabile

Assegnazione di un Mastro contabile al corrispondente rigo di bilancio



Responsabile della procedura Bilancio

GLAM0106

Funzione Che dato un Campo "Conto" Pic x(11)

Restituisce il campo "Rigo" PIC 9(7)



Sviluppatore SW

Database = DB2

Schema = NAACCT

Table = GLRACCO

Column = GL\_ACC\_NO

data type = int(07)

NS

DbA

Column = GL\_ACC\_NO  
data type = char(11)

# La Comunicazione tra le diverse culture...

....è una risorsa !



## Mastro Contabile

Identificativo del conto contabile, anche detto "contabilità"

Si compone di tre livelli

FFFFIIVVVV.

6

Che dato un  
Conto" Pic x(11)

Se il campo  
C 9(7)

.... MA ....

## Classificazione del Mastro Contabile

Assegnazione di un  
contabile al corrispettivo  
rigo di bilancio

= DB2

NAACCT

L-RACCO

GL\_ACC\_NO

= int(07)

NS

column = GL\_ACC\_NO

type = char(11)

# La Comunicazione tra le diverse culture...



....è molto difficile !

## Mastro Contabile

Identificativo di un conto, anche detto "capacità"

Si compone di un numero di 11 caratteri: FFFIIIVVVV.

A causa di...

- Differenze nel Linguaggio
- Mancanza di canali di comunicazione
- Mancanza della reciproca consapevolezza
- Organizzazione a funzioni separate

dato un campo "Piccolo" (11) campo (7)

....eppure parliamo di...

- Che fanno parte della stessa azienda
- Che possono avere maggiori reali dal buon funzionamento del sistema
- Che contribuiscono al medesimo processo produttivo
- Che lavorano con lo stesso sistema IT

32  
ACCT  
CCO  
ACC\_NO  
(07) NS  
= GL\_ACC\_NO  
e = char(11)

## Classificazione del Mastro Contabile

Assegnazione contabile al conto di riga di bilancio

# La Comunicazione tra le diverse culture...



## E' la missione del Business Glossary....

### Mastro Contabile

Identificativo del conto contabile anche detto "conto di contabilità"

Si compone di tre livelli :

FFFIIIVVVV.



Esperto di Contabilità

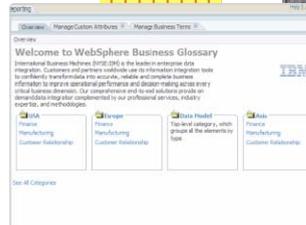
### Sviluppatore SW



### GLAM0106

Funzione Che dato un Campo "Conto" Pic x(11)

Restituisce il campo "Rigo" PIC 9(7)



### Classificazione del Mastro Contabile

Assegnazione di un Mastro contabile al corrispondente rigo di bilancio



Responsabile della procedura Bilancio

Database = DB2

Schema = NAACCT

Table = GLRACCO

Column = GL\_ACC\_NO

data type = int(07)

NS

Column = GL\_ACC\_NO  
data type = char(11)



DbA

# Il Business Glossary è uno strumento Web



WebSphere Data Integration Console - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://localhost:9080/> Go

**WebSphere** Data Integration Console

Home Administration **Glossary** Reporting Help | Log Out

**Navigation**

Browse Search History

Administration

- Edit Overview
- Manage Business Categories
- Manage Business Terms
- Manage Custom Attributes

Overview Manage Custom Attributes Manage Business Terms

Overview

## Welcome to WebSphere Business Glossary

International Business Machines (NYSE:IBM) is the leader in enterprise data integration. Customers and partners worldwide use its information integration tools to confidently transform data into accurate, reliable and complete business information to improve operational performance and decision-making across every critical business dimension. Our comprehensive end-to-end solutions provide on demand data integration complemented by our professional services, industry expertise, and methodologies.

**USA**  
Finance  
Manufacturing  
Customer Relationship

**Europe**  
Finance  
Manufacturing  
Customer Relationship

**Data Model**  
Top-level category, which groups all the elements by type

**Asia**  
Finance  
Manufacturing  
Customer Relationship

[See All Categories](#)

Done Local intranet

# Il Business Glossary è semplice ed intuitivo



Si compone di 3 soli tipi di oggetti:

Categorie

Termini

Collegamenti

The screenshot displays the WebSphere Data Integration Console interface. The browser window title is "WebSphere Data Integration Console - Microsoft Internet Explorer". The address bar shows "http://localhost:9080/". The main content area is titled "WebSphere Data Integration Console" and has tabs for "Home", "Administration", "Glossary", and "Reporting". The "Glossary" tab is active, showing an "Overview" page with the breadcrumb "Overview > All Categories". A left-hand navigation pane is titled "Navigation" and includes "Browse", "Search", and "History" buttons. Under "Administration", there are links for "Edit Overview", "Manage Business Categories", "Manage Business Terms", and "Manage Custom Attributes". The main area contains a grid of 16 category cards, each with a folder icon, a title, and a description:

- Transaction**: Metadata related to account transactions
- USA**: Finance, Manufacturing, Customer Relationship
- Finance**: General Ledger, Transaction
- General Ledger**: Metadata related to General Ledger accounts
- Manufacturing**: Metadata related to manufacturing
- Customer Re...**: Metadata related to CRM activities
- Europe**: Finance, Manufacturing, Customer Relationship
- Data Model**: Top-level category, which groups all the elements by type
- Finance**: Metadata related to Asian finances
- Manufacturing**: Metadata related to manufacturing
- Finance**: Metadata related to European finance
- Manufacturing**: Metadata related to manufacturing
- Customer Re...**: Metadata related to CRM activities
- Asia**: Finance, Manufacturing, Customer Relationship
- Customer Re...**: Metadata related to CRM activities
- AllTerms**: Special category containing all terms



Con il termine **tassonomia** ( dalle parole greche *taxis* = ordine e *nomos* = regole) ci si può riferire sia alla **classificazione gerarchica di concetti**, sia al principio stesso della classificazione.

Praticamente tutto (concetti, oggetti animati e non, luoghi ed eventi) può essere classificato seguendo uno schema tassonomico.

# Il Business Glossary è flessibile



## **Il Business Glossary può essere scritto dagli utenti**

- si possono gestire profili di “authoring” a diversi livelli
- si può assegnare uno “steward” responsabile di ciascun termine o categoria

## **Si possono riprendere termini e categorie da qualsiasi sorgente informatica**

- importazioni massive di termini
- importazione di strutture tassonomiche mappate in XML
- Importrazione tramite “bridge” da strumenti di modellazione di mercato

## **Si può estendere il modello dei metadati**

- annidando categorie e sottocategorie fino a creare la tassonomia voluta
- aggiungendo attributi al modello del Business Glossary
  - ✓ ai termini del Business Glossary
  - ✓ alle categorie e sottocategorie

# Il Business Glossary è integrato con IBM Information Server



## **L'utente Gestisce collegamenti tra termini del glossario**

- collegamenti logici
- sinonimi
- esempi d'uso comune

## **Business Glossary dispone di una ricerca semplice ed avanzata, che permette di ricercare contemporaneamente:**

- oggetti del Business Glossary
- oggetti tecnici (Metadati del Repository di IBM Information Server)
  - ✓ Job e loro componenti
  - ✓ Table definition (Tracciati)

## **Partendo dal risultato della ricerca, l'utente crea i collegamenti tra:**

- i termini di Business Glossary
- gli oggetti tecnici del repository di IBM Information Server

# Benefici di Business Glossary

- Consente un “authoring Web-based”, la condivisione e la gestione dei metadati di business
- Allinea le procedure IT con gli obiettivi di business
- Fornisce informazioni sul contesto del business agli asset informativi
- Stabilisce responsabilità e competenze sul dato



**Subject Matter Experts**   **Business Users**

Comprendi



## WebSphere Business Glossary

Consente la creazione e la gestione del business vocabulary, delle relazioni e dei collegamenti alle fonti fisiche del dato



# Il valore per l'azienda



## Il Business Glossary....

- Fornisce strumenti per una partecipazione attiva al business
- Aiuta a focalizzare il punto di vista del business
- Crea una più stretta collaborazione tra business ed IT
- Fornisce un framework per la condivisione di decisioni di business



### ... e permette di:

- Ridurre i tempi di analisi
- Allineare i processi IT con gli obiettivi di business
- Fornire informazioni puntuali e aggiornate sui dati in divenire
- Accelerare lo sviluppo di nuove regole di business

# I punti di forza di WebSphere Business Glossary



<b>Fornisce strumenti utili per il business</b>	<b><u>Benefici</u></b>
I business users possono creare termini e collaborare alla loro definizione, creare sinonimi ed esempi d'uso, costruire tassonomie che riflettono il punto di vista del business	Gli utenti possono contribuire attivamente alle specifiche di un progetto
Fornisce un ampio set di funzionalità per la gestione dei metadati di business, inclusi: reporting estensivo, ricerche, ricerche per sinonimi, analisi di impatto	Facilmente utilizzabile dai business users. L'accesso al ciclo di vita dei dati ne incoraggia il riutilizzo
<b>Facilita la collaborazione tra business &amp; IT</b>	
I termini e le tassonomie possono essere categorizzati, gestiti e condivisi tra differenti business units, gruppi, linee di prodotto o divisioni	Consente di approcciare i metadati da prospettive differenti, memorizzando tutti i diversi punti di vista
Tutti i metadati di business sono gestiti in un repository	Consente ai termini di business di essere integrati con l'analisi complessiva
Possibilità di creare annotazioni su termini e categorie.	Semplifica la comprensione del dato
Consente collegamenti tra termini di business e colonne o tabelle fisiche	Restringe i tempi di analisi
<b>Data Stewardship</b>	
Consente l'assegnazione di "data stewards" ad ogni termine	Miglior supporto per la governance



# Thank YOU

IBM Information  
On Demand 2008  
>>> Comes To You

*ALLA LUCE DELL'INFORMATION ON DEMAND*

Milano, 15 aprile 2008

