



Data Quality Management

Enrico Viola - CISA, membro ISO/JTC1/WG6, CMMI Appraiser – ECLAT srl
viola@eclat-web.com



1. Definizione ed obiettivi della DQ
2. Il modello di Data Quality
3. Approccio per processi
4. Data Quality e Basilea II



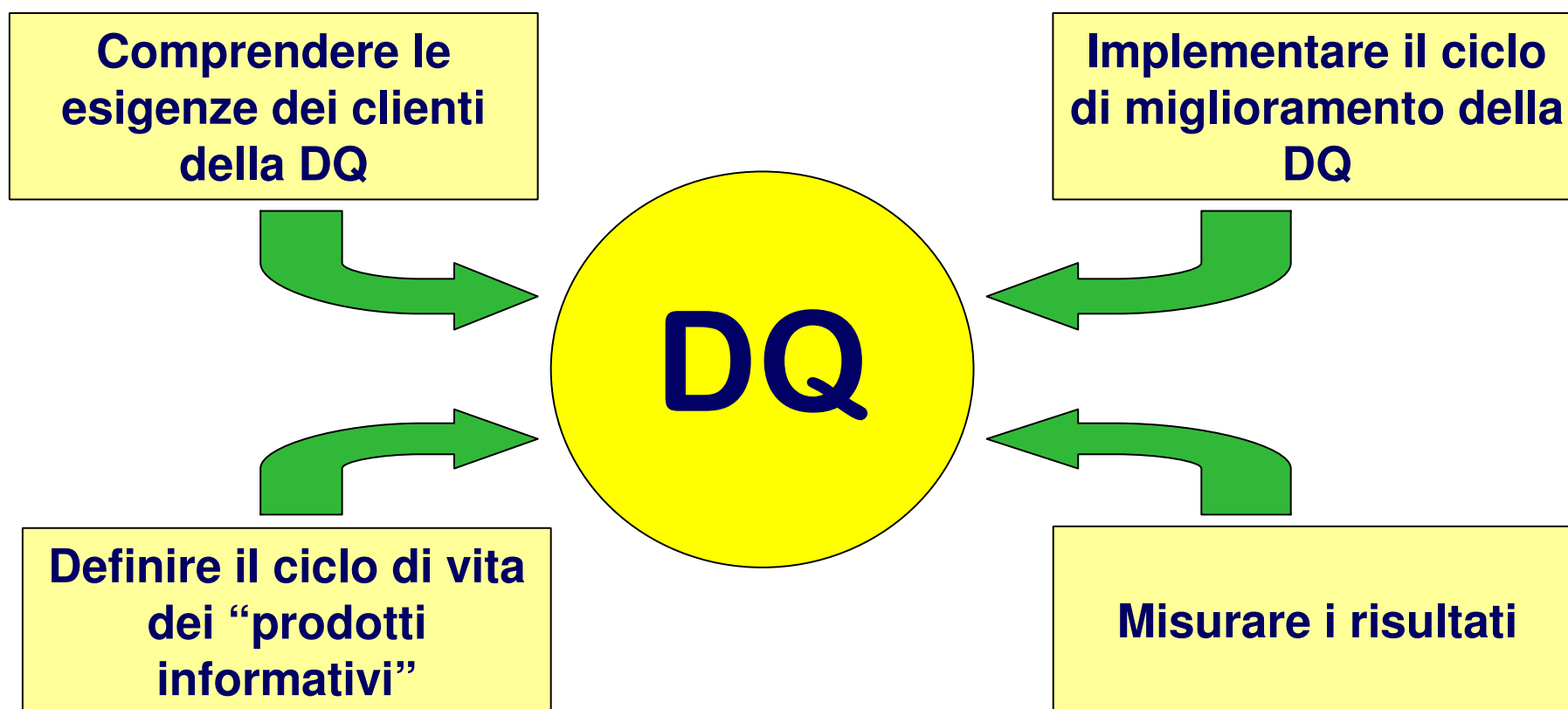
- ✓ Il concetto di “**QUALITA' DEI DATI**” è riferito alla capacità dei dati in possesso dell'organizzazione di **supportare le decisioni** necessarie a raggiungere i propri obiettivi
- ✓ La **QUALITA' DEI DATI** è uno dei **requisiti** necessari per l'efficienza di ogni organizzazione
- ✓ Ogni componente dell'organizzazione “usa” i dati, o partecipa al processo di “gestione” dei dati stessi, e contribuisce quindi alla **QUALITA' DEI DATI**



- ✓ E' un concetto “**multi-dimensionale**”: comprende aspetti **soggettivi** (es.: la convinzione degli utenti di disporre di informazioni affidabili e semplici da usare) ed aspetti **oggettivi** (es.: l'integrità dei dati e la riservatezza degli accessi)
- ✓ Un aspetto importante per l'organizzazione è il passaggio della DQ da variabile prodotta dal contesto a **prodotto di un processo strutturato come un progetto**



La Qualità dei Dati (DQ) è essa stessa uno dei
“prodotti” **dell' IT** dell'organizzazione





porre l'attenzione su

- ✓ I processi ed i soggetti coinvolti (utenti e fornitori) piuttosto che sui dati
- ✓ le dimensioni (rilevanti!) del concetto di DQ per l'organizzazione
- ✓ il processo di produzione delle informazioni
- ✓ la formazione dei managers
- ✓ I costi della “non qualità”

Obiettivo primario



Ridurre i costi della “NON QUALITÀ”

- a. costi conseguenti all'utilizzo di dati di scarsa qualità; esempi
 - ➔ *minore efficacia di decisioni prese per dati non aggiornati o non consistenti*
 - ➔ *opportunità perdute per valutazioni derivanti da dati non affidabili*
- b. spese per la ricerca e la correzione della “non qualità” di dati ed informazioni

I costi della “NON QUALITÀ”



Problems Due to Poor Data Quality

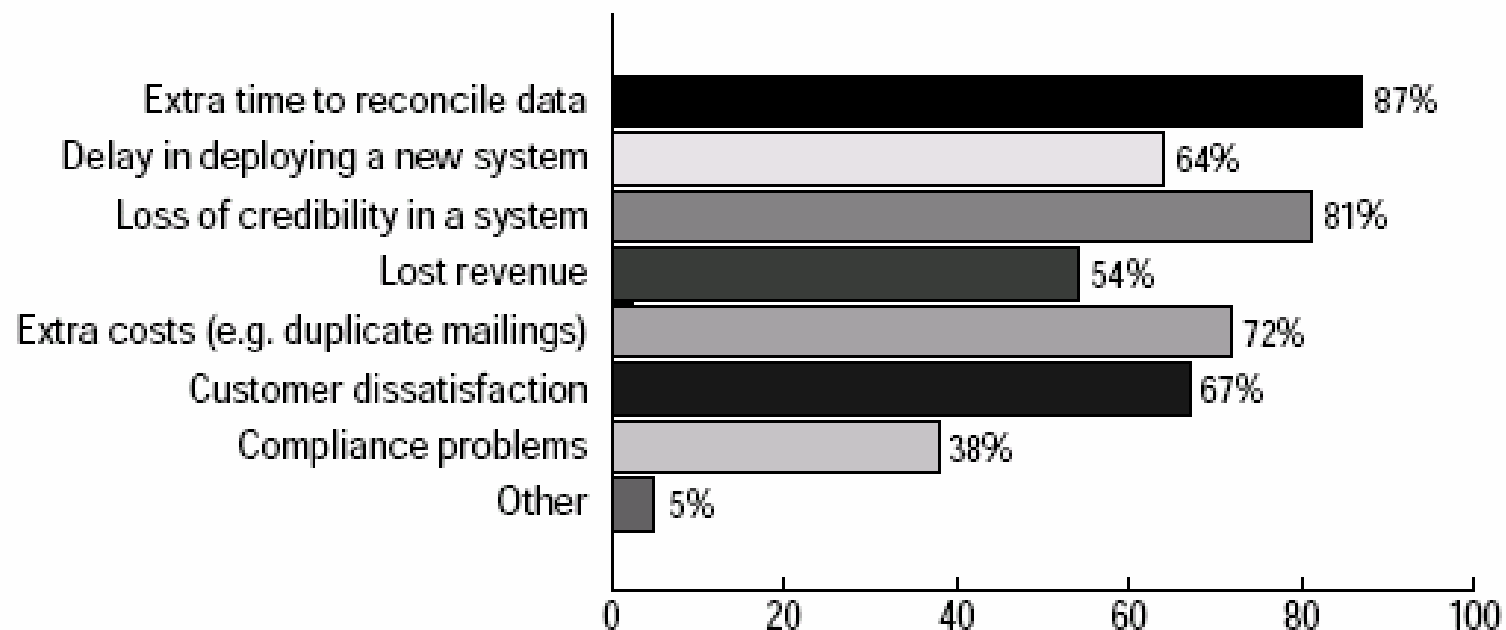


Illustration 7. Defective data causes a litany of problems. Based on 286 respondents who could select multiple answers. TDWI Data Quality Survey, December 2001.

I benefici della “QUALITA”



Benefits of High Quality Data

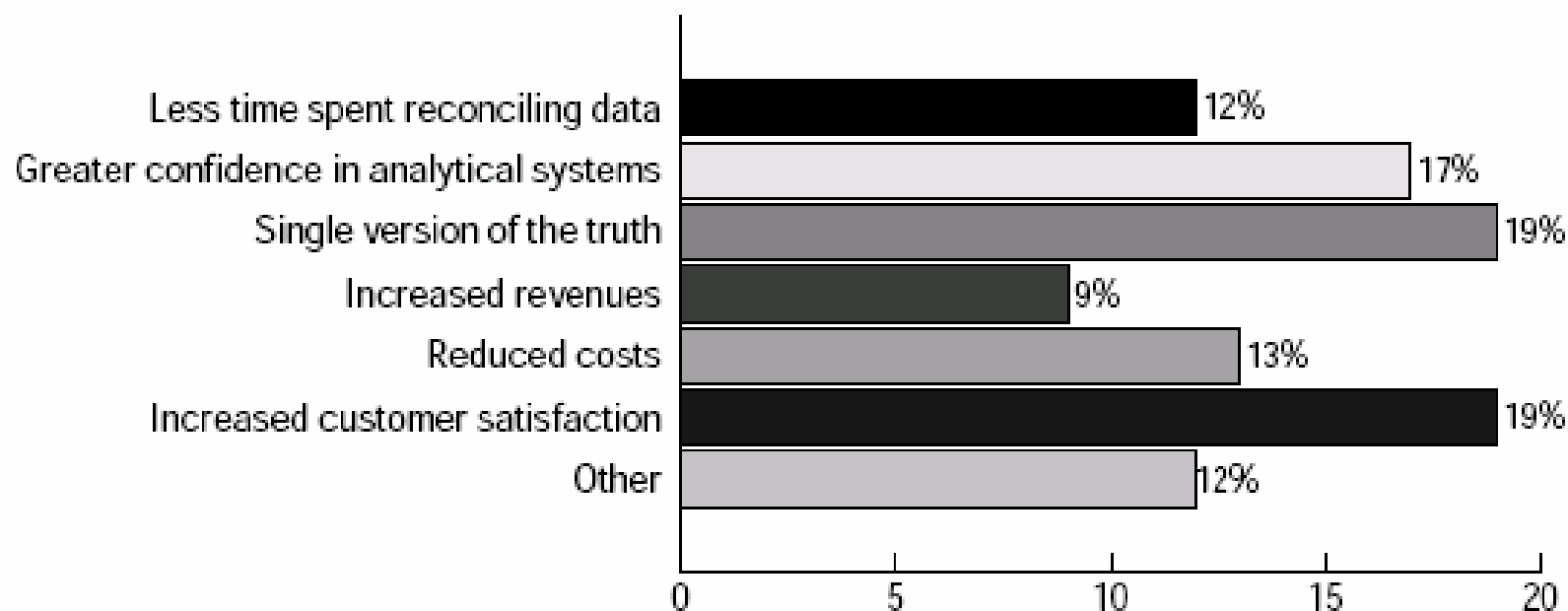


Illustration 8. There are many benefits to high quality data. Based on 304 responses, with respondents selecting one answer. TDWI survey on data quality, December 2001.



Sources of Data Quality Problems

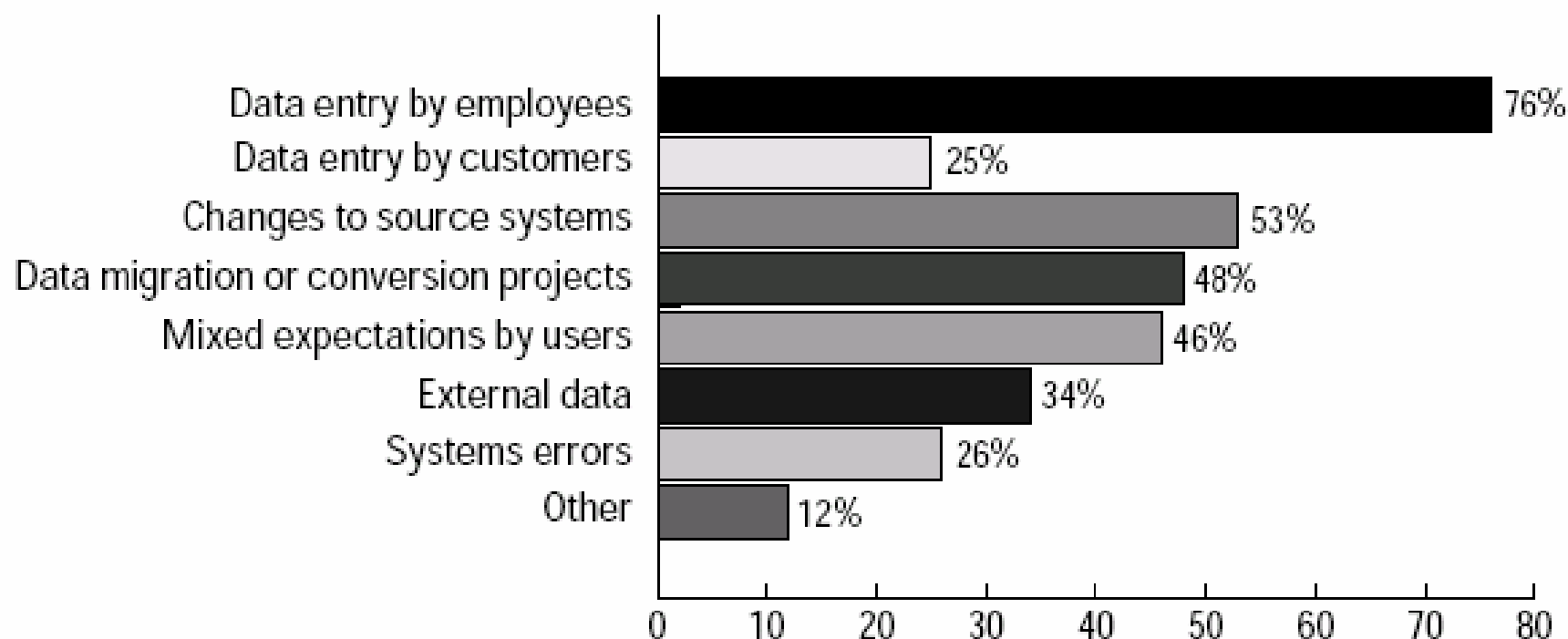


Illustration 9. Data entry and changes to source systems are the biggest causes of data quality problems for data warehousing managers.



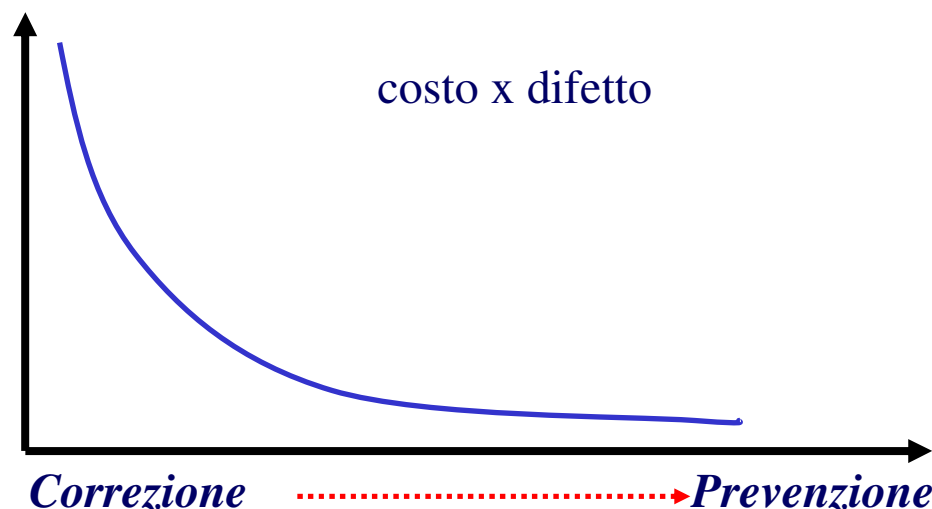
Diversi fattori rendono la Qualità dei Dati un elemento critico per le organizzazioni; tra questi:

- ✓ La necessità di tenere sotto controllo sistemi complessi ed interagenti (es.: ERP, CRM, Data Warehouse, servizi web)
- ✓ La brevità dei tempi a disposizione per l'assunzione di decisioni importanti per l'organizzazione
- ✓ La distribuzione dell'utenza potenziale dei dati/informazioni
- ✓ Il valore crescente del fattore “conoscenza” per l'organizzazione
- ✓ La dimensione e l'estensione crescenti delle basi di dati
- ✓ La diffusione di architetture “federate”



La DQ in pratica

Le soluzioni da adottare dovranno tenere conto delle reali esigenze dell'organizzazione e della struttura dei costi della Qualità dei Dati



Ogni soluzione di DQ sarà la combinazione di azioni tecniche a diversi livelli (es.: “real-time data integration” , data profiling, modifica di procedure e programmi esistenti) ed organizzative (es.: integrazione compiti DA / DBA).

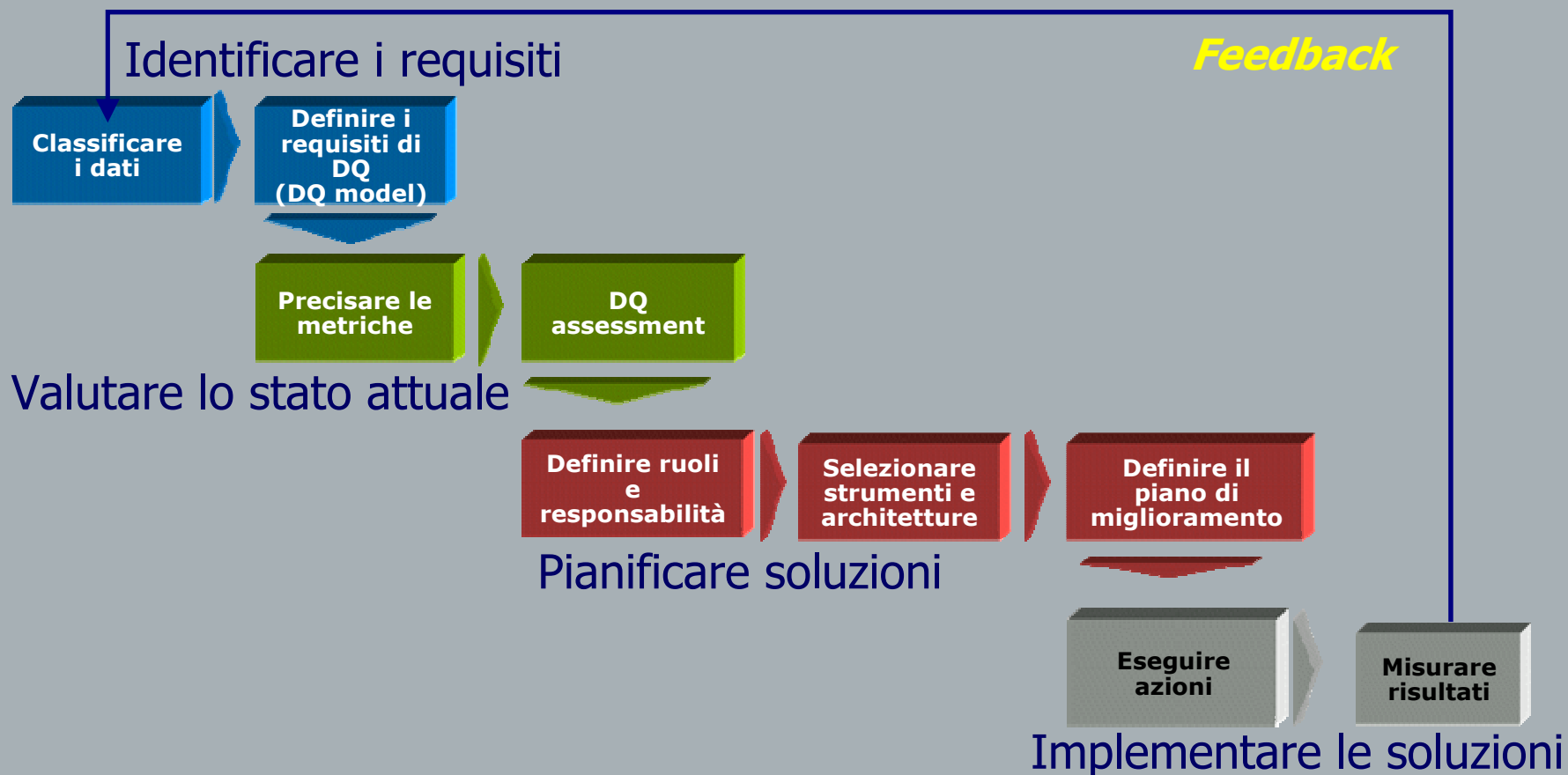


CARATTERISTICHE PRINCIPALI :

- ✓ Accoglie i principali contributi espressi dagli esperti di DQ Management (es: ISO, COBIT, TDQM, Tqdm)
 - ✓ Si ispira ai principi del TQM :
 - ➔ *Orientamento al cliente : ricerca del **VALORE***
 - ➔ ***Miglioramento continuo** : “un lungo viaggio inizia con un piccolo passo” (proverbio cinese)*
 - ➔ ***Misurabilità** (dei risultati e dei benefici)*
 - ✓ E' orientato ai processi : *una efficiente gestione della Qualità dei Dati elimina le cause dei problemi **correggendo i processi***
-



Schema semplificato dei processi di DQM



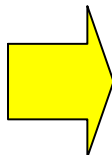


- ✓ I punti di vista da considerare sono tre:
 - La **PRODUZIONE** dei dati *(per prevenire le cause di errore)*
 - La **GESTIONE/MODIFICA** dei dati *(per il mantenimento/miglioramento della DQ)*
 - La **FRUIZIONE** dei dati *(UTENTI della DQ, per focalizzarsi sulle esigenze primarie)*
- ✓ il processo di gestione della DQ coinvolge in modo importante anche i partners (i clienti nel caso di Basilea II)
- ✓ ogni utente è responsabile dei dati trattati o trasferiti al “sistema informativo”
- ✓ I processi costituiscono il punto di riferimento per la determinazione delle regole della DQ



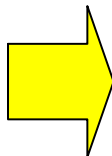
3 obiettivi principali

1. apportare VALORE
all'organizzazione



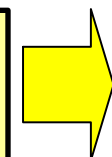
- **Approccio per processi**
- Focus sul cliente (es. *Formazione, comunicazione, priorità*)

2. Miglioramento continuo



- caratteristiche, metriche e strumenti di DQ sono in continua evoluzione
- benefici quantificabili

3. Valorizzare il processo di
Data Quality Management

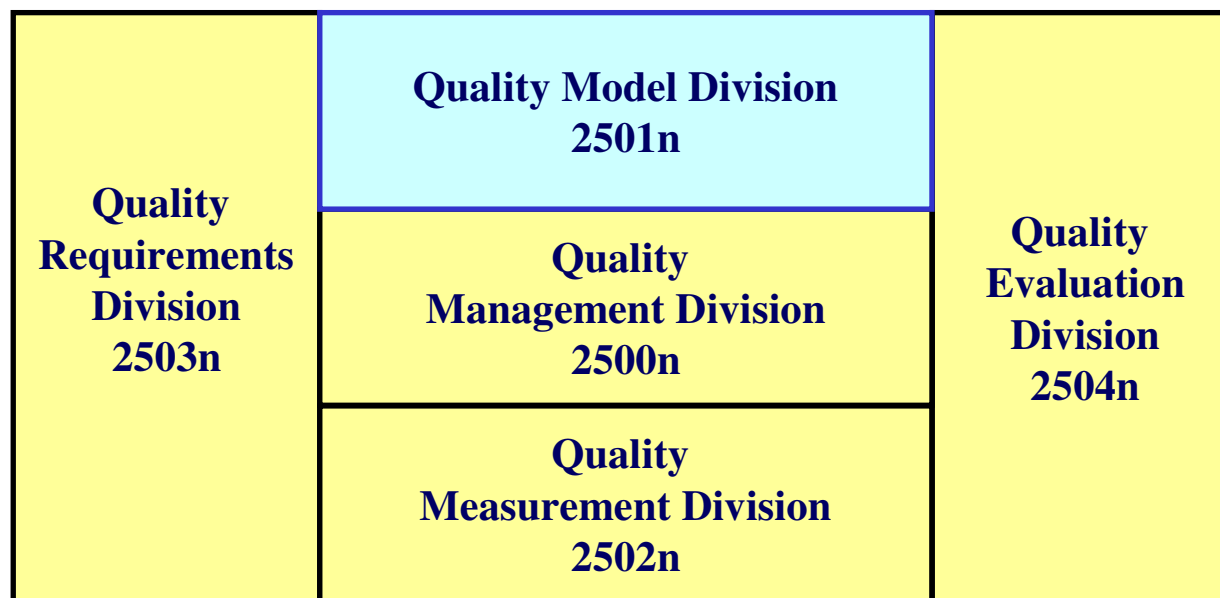


- supporto di tecniche e strumenti adeguati
- qualificazione professionale



Il progetto ISO : SQuaRE

**la Qualità dei Dati è inserita nel progetto ISO
SQuaRE *Software Product Quality Requirements
and Evaluation*, condotto nell'ambito del JTC1/WG6**





Il Progetto ISO SQuaRE

➔ **ISO/IEC 2500n - Quality Management Division.**

Le norme che fanno parte di questa “divisione” definiscono tutti i modelli ed i termini comuni a tutto il progetto SQuaRE. E' presente anche una guida ai documenti SQuaRE ed una serie di suggerimenti pratici di alto livello sull'applicazione degli standards appropriati a casi specifici.

Questa divisione fornisce inoltre requisiti e consigli per una funzione di supporto che è responsabile della specifica e valutazione dei requisiti dei prodotti software.

➔ **ISO/IEC 2501n - Quality Model Division.**

Le norme che fanno parte di questa “divisione” presentano un modello di qualità dettagliato che comprende le caratteristiche di qualità interna, esterna ed in uso. Inoltre le caratteristiche interne ed esterne sono scomposte in sotto-caratteristiche. E' presente anche una guida pratica per l'uso del modello.



Il Progetto ISO SQuaRE

➔ **ISO/IEC 2502n - Quality Measurement Division.**

Le norme che fanno parte di questa “divisione” includono un modello di riferimento per la misura della qualità di un prodotto software, delle definizioni matematiche delle misure di qualità, ed una guida pratica per la loro applicazione. Le misure esposte si riferiscono alla qualità del software interno, esterno ed in uso.

➔ **ISO/IEC 2503n - Quality Requirements Division.**

Le norme che fanno parte di questa “divisione” supportano la specifica di requisiti di qualità, che possono essere usati per la specifica dei requisiti di un prodotto software da sviluppare o come input di un processo di valutazione. il processo di definizione dei requisiti corrisponde ai processi definiti dalla norma ISO/IEC 15288 – Information Technology - Life Cycle Management - System Life Cycle Processes.

Il Progetto ISO SQuaRE

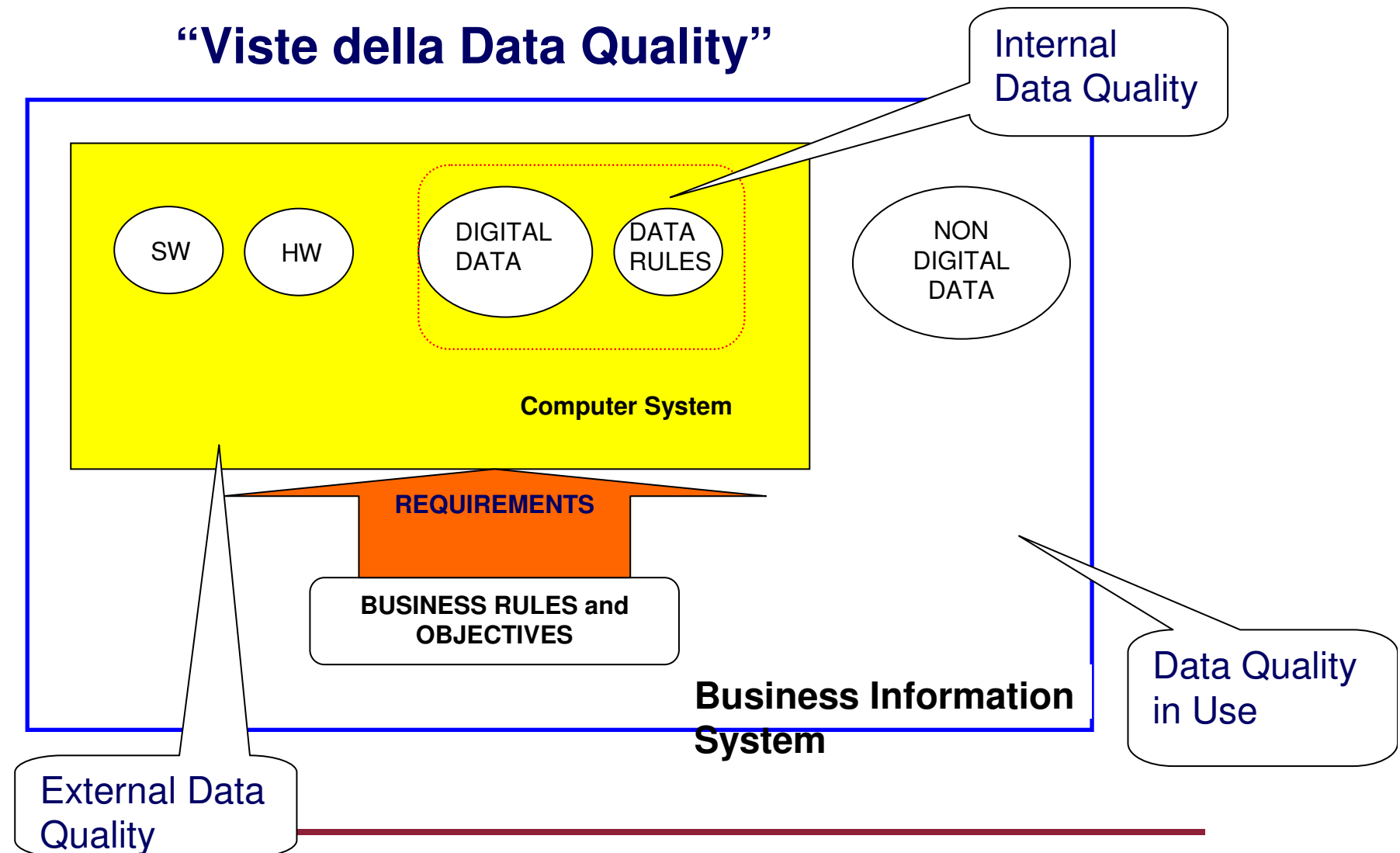


➡ ISO/IEC 2504n - Quality Evaluation Division.

Le norme che fanno parte di questa “divisione” forniscono requisiti, raccomandazioni e linee guida per la valutazione del prodotto software, eseguita da valutatori, acquirenti o sviluppatori.



“Viste della Data Quality”



Il modello ISO/IEC 25012



QUALITA' DEI DATI INTERNA / ESTERNA

Caratteristica		Caratteristica	
Funzionalità	Consistenza	Usabilità	Comprensibilità
	Frequenza di aggiornamento		Gestibilità
	Completezza		Attrattiva
	Precisione		Conformità (usabilità)
	Accuratezza	Efficienza	Tempo di risposta
	Interoperabilità		Utiizzo di risorse
	Sicurezza		Conformità (efficienza)
	Conformità (funzionalità)		Analizzabilità
Affidabilità	Disponibilità	Manutenibilità	Modificabilità
	Ripristinabilità		Conformità(manutenibilità)
	Conformità (affidabilità)	Portabilità	Adattabilità
			Conformità (portabilità)



Il modello ISO/IEC 25012

QUALITA' IN USO		
Caratteristica		
Efficacia	Aggiornamento / Tempestività	
	Completezza informativa	
	Conformità (efficacia)	
Produttività	Rilevanza	
	Interpretabilità	
	Conformità (produttività)	
Safety	Safety relevance	
	Conformità (safety)	
Soddisfazione	Credibilità	
	Accessibilità	
	Conformità	

ISO/IEC 25012 – Data



“Viste” della Qualità dei Dati: qualità INTERNA **Quality Model**

Rappresenta la capacità dei dati di soddisfare le esigenze dell'utente, in condizioni date. Le caratteristiche di qualità “interna” si riferiscono ai dati “di per sé”; esse forniscono i criteri per garantire e verificare la qualità di:

- valore
- tipo e lunghezza
- definizioni (incluso i metadati)
- regole (es.: regole di integrità/consistenza)
- relazioni tra dati

avendo a riferimento la documentazione, gli archivi fisici ed il mondo reale.

Esempio: la “consistenza” si riferisce ai dati indipendentemente dagli aspetti legati all'hardware o al software.

Gli utenti coinvolti sono tutti coloro che partecipano alle fasi di “preparazione” e di “acquisizione” dei dati.



ISO/IEC 25012 – Data Quality

Model

“Viste” della Qualità dei Dati: qualità ESTERNA

Si riferisce ad alcune proprietà di componenti del “sistema” (es: Hardware, Data Base Management Systems, Software) applicate ai dati in esso contenuti.

Esempi: la Sicurezza delle informazioni dipende dall’implementazione di funzionalità tipiche del sistema, che possono essere applicate con diversa intensità ed estensione ai dati.

NOTE :

- (1) La qualità dei dati “interna” non fornisce indicazioni circa la qualità dei dati “esterna”;
- (2) le metriche per la misurazione possono essere differenti nel caso in cui una stessa caratteristica abbia rilievo sia per la qualità “interna” che per quella “esterna”.



ISO/IEC 25012 – Data Quality

Model

“Viste” della Qualità dei Dati: qualità IN USO

rappresenta la capacità dei dati di mettere in grado gli utenti di raggiungere i propri obiettivi attraverso il loro utilizzo in un determinato contesto.

Gli utenti della qualità in uso possono essere sia i progettisti di applicazioni software sia gli utenti finali del sistema informativo;

La qualità dei dati “in uso” ha l’obiettivo di definire le caratteristiche di qualità dei dati che esprimono il punto di vista *soggettivo* dell’utente a proposito del suo grado di soddisfazione circa la rispondenza dei dati alle sue esigenze operative

Caratteristiche: qualità interna /



Caratteristica	Sotto-caratteristica	Definizione
Funzionalità	Consistenza	Assenza di contraddizioni tra i dati di uno stesso record o di archivi diversi. Ad esempio :all'interno dell'entità "impiegati" la data di nascita non può essere maggiore della data di assunzione. Lo stesso se i dati sono contenuti in archivi diversi.
	Frequenza di aggiornamento	I dati sono aggiornati in relazione alle necessità. E' una dimensione critica per i dati "volatili" (es. la velocità del vento o la temperatura climatica) Es.: Un posto passeggeri su un aereo non deve essere più disponibile dopo che è stato assegnato a qualcuno
	Completezza	Tutti i valori necessari sono stati assegnati e memorizzati nel Computer system. Può essere riferita al mondo reale (in questo caso è il rapporto tra i dati degli archivi e quelli del mondo reale) o al grado di riempimento degli archivi (in questo caso il rapporto è tra i "campi riempiti" e il numero dei campi) Es.: con riferimento all'entità "dipendente" di una organizzazione, tutte le istanze ad essa relative devono contenere gli attributi richiesti dagli utenti (in sede di definizione dei requisiti delle applicazioni software relative)
	Precisione	E' la capacità di un valore assegnato ad un attributo di fornire il grado di informazione necessaria in un determinato contesto. Ad esempio : per rappresentare la durata di una maratona, l'unità di tempo minima è i "secondi", mentre per rappresentare la durata di una corsa sui 100 metri l'unità di tempo minima è i "millisecondi"

Caratteristiche qualità interna /



Caratteristica	Sotto-caratteristica	Definizione
Funzionalità	Accuratezza (semantica e sintattica)	<p>E' il grado di conformità di un valore acquisito, misurato e calcolato, al suo attuale valore. E' caratterizzata principalmente da due aspetti</p> <p>Accuratezza sintattica E' la precisione del valore di un dato rispetto ad un dominio di valori considerato sintatticamente corretto. Ad esempio, un basso grado di accuratezza sintattica si ha se il nome Maria è riportato come Marja</p> <p>Accuratezza semantica E' la precisione del valore di un dato rispetto ad un dominio di valori considerato semanticamente corretto. Ad esempio, un basso grado di accuratezza semantica si ha se il nome Maria è riportato come Francesca. Francesca è sintatticamente accurato, perché è compreso nel dominio di riferimento, ma è un nome diverso.</p>
	Interoperabilità	E' possibile accedere a scambiare i dati tra piattaforme tecnologiche e sistemi differenti.
	Sicurezza	Solo gli utenti autorizzati possono accedere ai dati
	Conformità alle norme	E' la proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la funzionalità.

Caratteristiche: qualità interna / esterna (3/4)



Affidabilità	Disponibilità	i dati sono sempre disponibili . Nota: un caso particolare di disponibilità è l'accesso concorrente (in lettura e scrittura) da parte di più di un utente e/o applicazione.
	Ripristinabilità	E' la proprietà dei dati di mantenere l'integrità fisica e logica anche in caso di guasti. Lo standard si riferisce alla qualità dei dati, non del sistema: la ripristinabilità si riferisce pertanto ai singoli dati "ripristinabili". (ad es. attraverso meccanismi come: commit/synchpoint, rollback e backup-recovery).
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa l'affidabilità.
Usabilità	Comprensibilità	I dati sono espressi con linguaggio e simboli appropriati. Le definizioni sono espresse con chiarezza. Alcune informazioni circa la comprensibilità sono fornite dai metadati. Esempio: per rappresentare uno Stato, l'acronimo standard è più comprensibile di un codice numerico
	Gestibilità	I dati sono facilmente gestibili da un punto di vista funzionale. NOTA :Il campo che rappresenta un importo di costo data deve essere memorizzato come "numerico" e non come "stringa di caratteri", per permettere operazioni algebriche
	Attrattiva	I dati devono attrarre l'attenzione dell'utente. Esempio: per descrivere un villaggio turistico un dato registrato come immagine ha maggiore attrattiva di una descrizione testuale
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa l'usabilità

Caratteristiche: qualità interna / esterna (4/4)



Efficienza	Richiesta di risorse	E' la possibilità di poter accedere e memorizzare i dati utilizzando una quantità di risorse appropriate con riferimento a condizioni date Esempio: usare più spazio del necessario per memorizzare un dato può provocare spreco di spazio, di memoria e di CPU
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa l'efficienza.
Modificabilità	Analizzabilità	rappresenta la possibilità di analizzare e testare i dati
	Modificabilità	E' la proprietà di poter modificare il formato, lunghezza e valore assegnato ai dati
	Conformità	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la modificabilità.
Portabilità	Adattabilità	Rappresenta la possibilità di spostare i dati da una piattaforma tecnologica ad un'altra Include la possibilità di installare e replicare i dati nella piattaforma di destinazione. Si riferisce in modo particolare ad insiemi coerenti di dati (ad esempio quelli richiesti da una specifica applicazione software)
	Conformità	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la portabilità.

Caratteristiche: qualità in uso



(1/2)

Efficacia	Aggiornamento / Tempestività	E' il grado di aggiornamento ritenuto sufficiente per le esigenze dell'utente Esempio: per poter frequentare un corso uno studente ha bisogno di conoscere l'esatto calendario delle lezioni prima della data di inizio
	Completezza informativa	E' la completezza dei dati dal punto di vista dell'utente e delle sue esigenze da un punto di vista quantitativo. Comprende anche la capacità dei dati a rappresentare il contesto (realtà) osservato dall'utente Esempio : i dati sono caratterizzati da completezza informativa se sono rappresentati tutti gli attributi necessari e ci sono valori assegnati agli attributi quando sono necessari
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa l'efficacia.
Produttività	Rilevanza	Il grado di utilità del dato in ragione delle esigenze I dati sono rilevanti per un utente quando rispondono all'ambito di utilizzo ed agli obiettivi dei bisogni informativi
	Interpretabilità	E' il grado di comprensibilità dei dati NOTA : I Metadati possono facilitare la comprensione dei dati
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la produttività.

Caratteristiche: qualità in uso



(1/2)

Safety	Safety	La capacità dei dati di garantire un livello accettabile di rischio di danno alle persone, all'ambiente ed al "business" in uno specifico contesto. Esempio: i dati relativi al gruppo sanguigno di un individuo hanno rilievo per la safety
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa l'incolumità delle persone, dell'ambiente e del business
Soddisfazione	Credibilità	I dati sono ritenuti veri e credibili.
	Accessibilità	L'accesso ai dati è consentito, con particolare riferimento alle persone che hanno bisogno di particolari tecnologie o configurazioni a motivo di alcune disabilità.
	Conformità alle norme	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la soddisfazione.

Le dimensioni della DQ



The academics' view of information quality

	Intrinsic IQ	Contextual IQ	Representational IQ	Accessibility IQ
Wang and Strong [39]	Accuracy, believability, reputation, objectivity	Value-added, relevance, completeness, timeliness, appropriate amount	Understandability, interpretability, concise representation, consistent representation	Accessibility, ease of operations, security
Zmud [41]	Accurate, factual	Quantity, reliable/timely	Arrangement, readable, reasonable	
Jarke and Vassiliou [16]	Believability, accuracy, credibility, consistency, completeness	Relevance, usage, timeliness, source currency, data warehouse currency, non-volatility	Interpretability, syntax, version control, semantics, aliases, origin	Accessibility, system availability, transaction availability, privileges
Delone and McLean [11]	Accuracy, precision, reliability, freedom from bias	Importance, relevance, usefulness, informativeness, content, sufficiency, completeness, currency, timeliness	Understandability, readability, clarity, format, appearance, conciseness, uniqueness, comparability	Usableness, quantitateness, convenience of access ^a
Goodhue [14]	Accuracy, reliability	Currency, level of detail	Compatibility, meaning, presentation, lack of confusion	Accessibility, assistance, ease of use (of h/w, s/w), locatability
Ballou and Pazer [4]	Accuracy, consistency	Completeness, timeliness		
Wand and Wang [37]	Correctness, unambiguous	Completeness	Meaningfulness	

Fonte : Information & Management 40 (2002)

Qualità dei Dati e Basilea II



Le **principali carenze** individuate dal Comitato nell'uso dei modelli per il rischio di credito come base di calcolo dei requisiti patrimoniali minimi riguardano la **qualità dei dati** e la capacità delle banche e delle autorità di vigilanza di validare i risultati dei modelli. I sistemi di rating interni sono un elemento fondamentale in molti modelli per il rischio di credito e pertanto questi aspetti – qualità dei dati e validazione – sono importanti tanto per i sistemi IRB quanto per la modellizzazione del rischio di credito. Tuttavia, il Comitato ritiene che nell'ambito del sistema IRB sia possibile ovviare a tali carenze stabilendo rigorosi requisiti minimi che le banche devono soddisfare nel definire input e output dei loro sistemi ed escludendo a questo stadio la valutazione da parte delle banche stesse di effetti di portafoglio quali la concentrazione e la diversificazione

(Comitato di Basilea per la vigilanza bancaria:

Presentazione del Nuovo Accordo di Basilea sui requisiti patrimoniali, gennaio 2001)



La Qualità dei Dati non è un requisito esclusivo di Basilea II

Alcuni esempi :

- ✓ *le norme di sicurezza BS7799 (Confidentiality, Integrity, Availability)*
- ✓ *la norma della Banca d'Italia relativa al piano di continuità operativa*
- ✓ *la legge sulla privacy (Dlgs 196/2003)*
- ✓ *le norme interne relative al Data Management (es.: COBIT)*
- ✓ *i requisiti delle singole applicazioni*

Qualità dei Dati e ISO17799



Information security is characterized here as the preservation of:

- a) **confidentiality**: ensuring that information is accessible only to those authorized to have access;
- b) **integrity**: safeguarding the accuracy and completeness of information and processing methods;
- c) **availability**: ensuring that authorized users have access to information and associated assets when required.

Information security is achieved by implementing a suitable set of controls, which could be policies, practices, procedures, organizational structures and software functions. These controls need to be established to ensure that the specific security objectives of the organization are met.

ISO/IEC 17799:2000, introduction



Il modello di Qualità dei Dati

Deve tenere conto di tutti i requisiti dell'organizzazione

Es.: Banca dati per il calcolo del rischio di credito

	Basilea II	BCP	BS7799	Privacy
Consistenza	X				
Aggiornamento	X	X			
Accuratezza	X		X	X	
.....					

Nota : le “X” rappresentano :

- descrizione del requisito
- obiettivo dell'organizzazione espresso in forma quantitativa (indicatori)

Il modello di Qualità dei Dati



- ✓ Il modello di qualità dei dati ISO (integrato ove necessario, ad esempio con COBIT) può costituire la base per definire il modello di qualità dei dati dell'organizzazione.
 - ✓ Esso contribuisce alle fasi di:
 - a) *definizione del modello di qualità dei dati dell'organizzazione*
 - b) *supporto ai processi di audit (sia interno che degli organi di vigilanza) per la verifica delle conformità*
 - c) *definizione dei requisiti per i dati esterni per l'integrazione con il sistema informativo della banca*
 - d) *realizzazione di sistemi software che rispondano ai requisiti funzionali della banca, tra cui quelli dettati dalle norme di Basilea*
-

Qualità dei Dati e Basilea II



Esempi tratti dal documento :

Guidelines on the implementation, validation and assessment of Advanced Measurement (AMA) and Internal Ratings Based (IRB) Approaches

CEBS-Committee of European Banking Supervisors – 11 luglio 2005



a) definizione del modello di qualità dei dati dell'organizzazione

2. Cooperation procedures, approval and post-approval process

2.2.1.1. Minimum Content

Documentation of rating systems

50. For an application to use the IRB Approach, it is essential that the documentation of rating systems include at least:

- ➔ A list of all internal documents held by the applicant that it considers relevant to the application, including a brief description of their contents.
- ➔ A map of the models that the institution will use for each of its portfolios.
- ➔ A general description of each of the models. **This can include a description of the types of data used, the definitions, classifications and the methodologies used, and some quantitative and qualitative assessments.**



a) definizione del modello di qualità dei dati dell'organizzazione

3.4.1. Data accuracy, completeness and appropriateness

295. Annex VII, Part 4, Paragraph 31 of the CRD requires the data used as inputs into the model to be accurate, complete, and appropriate. This refers to the use of a statistical model and other methods to assign exposures to obligors or facilities grades or pools. **Within this context**, these terms should be interpreted to have the following meaning:

- 296. 'Accurate'** refers to the degree of confidence that can be placed in the data inputs. Data must be sufficiently accurate to avoid material distortion of the outcome.
- 297. 'Complete'** means that databases provide comprehensive information for the institution (i.e., data for all relevant business lines and all relevant variables). While missing data for some fields or records may be inevitable, institutions should attempt to minimise their occurrence and aim to reduce them over time.
- 298. 'Appropriate'** means that data do not contain biases which make them unfit for purpose



**Proposta di DIRETTIVE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO
Rifusione della direttiva 2000/12/CE del Parlamento europeo e del Consiglio, del
20 marzo 2000, relativa all'accesso all'attività degli enti creditizi e al suo esercizio e della
direttiva 93/6/CEE del Consiglio, del 15 marzo 1993, relativa all'adeguatezza
patrimoniale delle imprese di investimento e degli enti creditizi**

.....

1.4. Impiego dei modelli

***31. Se un ente creditizio utilizza modelli statistici ed altri metodi automatici per
l'assegnazione delle esposizioni a gradi di merito o ad aggregati relativi a debitori o ad
operazioni, allora:***

.....

- b) l'ente creditizio dispone di un processo per vagliare i dati immessi nel modello di previsione che contempli una valutazione dell'accuratezza, completezza e pertinenza dei dati;***
- c) l'ente creditizio dimostra che i dati impiegati per costruire il modello sono rappresentativi della sua effettiva popolazione di debitori e di esposizioni;***
- d)***



a) **definizione del modello di qualità dei dati dell'organizzazione**

3.4.2. Data quality standards and consistency with accounting data

304. Institutions should define their own standards for ensuring data quality and should strive to improve these standards over time. They should measure their performance against these standards.

Institutions should work on an ongoing basis to ensure that their data is of high enough quality to support their risk management processes and the calculation of their capital requirements. This could include reviewing the structure of input data to identify outliers or implausible values, changes from previous periods, and the amount of missing data. **The review should also indicate whether the integrity of data is being maintained.**



a) definizione del modello di qualità dei dati dell'organizzazione

308. The following types of documentation are essential:

- ✓ **Data policy and statement of responsibility:** Institutions should establish an explicit data policy. They are responsible for ensuring the quality of their data and should be able to convince supervisors that they are meeting fit for purpose standards.
- ✓ **Data directory:** Institutions should have clear data directories (dictionaries) that provide definitions of data items.
- ✓ **Database descriptions:** Documentation for databases should allow institutions to provide supervisors with the information needed to determine the soundness of the databases. For example:
 - A general description of the databases (for example, information on the relational database model, including tables, keys, triggers, stored procedures; performance data such as overall maximal size; security information such as owner, users with read and write access, and maintenance responsibilities).
 - The source of the data.
 - The processes used to obtain and load the data.
 - The filters used to create and debug the database (e.g., maximum and minimum values and treatment of missing values).
 - Controls on access, consistency, etc.
 - A specific description of the variables included.



b) supporto ai processi di audit (sia interno che degli organi di vigilanza) per la verifica delle conformità

2.2.2. Supervisor's Assessment

76. The issues that supervisors are expected to assess can be broken down into five phases:

1. Methodology and documentation.
2. **Data quality.**
3. Quantitative procedures.
4. Qualitative procedures.
5. Technological environment.

77. In assessing IRB applications, the issues are:

1.
 2. **Data quality: Supervisors will assess the quality of data and databases** being used for the development of the rating systems, in the rating assignment process, and in the estimation of risk parameters, along with any other databases needed to calculate minimum regulatory capital. This is to ensure the soundness of calibration and of the capital calculation. (The latter holds true also for the other four items mentioned here.)
-



b) supporto ai processi di audit (sia interno che degli organi di vigilanza) per la verifica delle conformità

Audit review

- 303. Internal Audit of data quality should include at least the following: an annual review of controls, periodic sampling of data, and review of system reconciliations. For external data and pooled data, this work should be done to the fullest extent possible.**
- 306. There should be minimum checks including a periodic review by an independent party to confirm that data are accurate, complete, and appropriate. For example, data quality could be reviewed by replicating the preparation of data (including collection and transformation) and the outputs of models, using the same databases and algorithms used by the institution. This could be done on a sample basis. Supervisors could assess specific parts of this process to verify the conclusions reached by the independent reviewer.**



c) definizione dei requisiti per i dati esterni per l'integrazione con il sistema informativo della banca

3. Supervisor's assessment of the application concerning the minimum requirements of the CRD – Credit Risk

3.3.5. External vendor models

Linkage to internal information used in the rating process

287. The institution has to know what information (data) is processed in the external model and **how this information is linked** to information that is processed in house:
for example, if the vendor's definition of input factors such as sales and debt are consistent with those used in house.

289. The institution has to prove that the data used to build the external model is **representative of the actual portfolio**. If the model is used only for assigning exposures (i.e., for ranking borrowers in relative terms) the proof of consistency with the CRD's definition of default (Annex VII, Part 4, Paragraph 44) can be omitted

Gli strumenti per la Qualità dei Dati



Esempio di classificazione degli strumenti

Strumenti di analisi della QD e di indagine

Funzionalità : Validazione metadati
Misura della Qualità
Pattern Analysis (Business Rules Discovery)
Relationship Discovery & validation
Data Profiling e data filtering
Monitoring e Tracing DQ

Strumenti di correzione dei dati

Funzionalità : Data cleaning
Linking & matching
Data migration (standardizzazione, trasformazione, correzione,...)

Strumenti di Gestione

Funzionalità : Disegno Basi di dati
Reengineering
Gestione metadati
Conduzione assessment e Policy compliance
Integrazione dati ed applicazione
Prevenzione per la DQ



Conclusione

Prestare la massima attenzione (nel processo e nella realizzazione del prodotto) alla qualità dei dati, alle informazioni percepite dall'utente, all'ergonomia del software, tenendo conto della gerarchia di oggetti trattati che incorporano una crescente quantità di informazione:

- Dati (rappresentazione atomica interpretabile) e regole d'uso
- Informazione interscambiabile (es. XML)
- Documenti, studi e analisi, interfaccia web
- Cataloghi, dizionari, banche dati integrate e rete internet (conoscenza digitale)
- Conoscenza (KM)



- English, Larry P. Improving Data Warehouse and Business Information Quality. New York: John Wiley & Sons, 1999.
- K. Huang, Y. Lee, R. Wang, Quality Information and Knowledge, Prentice Hall, Upper Saddle River, NJ, 1999.
- Lee Y.W., Pipino L.L., Wang R.Y., Data Quality Assessment, Communications of the ACM, Volume 45, No. 4, April 2002.
- Mecella M., Scannapieco M., Virgillito A., Baldoni R., Catarci T., Batini C.: Managing Data Quality in Cooperative Information Systems. In Proceedings of the Tenth International Conference on Cooperative
- Domenico Natale, Monica Scannapieco, Paola Angeletti, Tiziana Catarci, Gianluigi Raiss : Qualità dei dati e Standard ISO/IEC 9126: Analisi Critica ed Esperienze nella Pubblica Amministrazione Italiana
- Redman T. C., Data Quality for the Information Age. Artech House, 1996.
- Redman T. C., 'Data Quality: A field Guide, Digital Press', Jan 15 2001.



grazie per l'attenzione!

Enrico Viola
viola@eclat-web.com