



0101010111100001101010111010101110101001010110010101001010010011110100001010101000110101101001  
00011101011101010100101010101010011100011010101010101010001111010101010010010100011101010101

# ISO/IEC 25012 : DATA QUALITY MODEL

Relatore: Domenico Natale

010101011110000110101011101010111010100101011001010100101001001111010000101010100011010110100  
00011101011101010100101010101010011100011010101010101010001111010101010010010100011101010101







# Contesto generale

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101010101010101010101001110001101010101010101010001111010101010010010100011101010101

- “Ogni generazione sarà più digitale di quella che l’ha preceduta”  
(Nicholas Negroponte, Essere digitale, 1995)
- Utenti internet a livello mondiale:
  - ✓ 2000: 195 milioni
  - ✓ 2002: 495 milioni
  - ✓ 2005: 965 milioni

01010101111000011010101110101011101010010101100101010010100100111101000010101010100011010110100  
0001110101110101010010101010101010011100011010101010101010001111010101010010010100011101010101





# Il Progetto ISO SQuaRE

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
000111010111010101010101010101010001110001101010101010101010001111010101010010010100011101010101

## ➔ **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.

## ➔ **ISO/IEC 2504n - Quality Evaluation Division.**

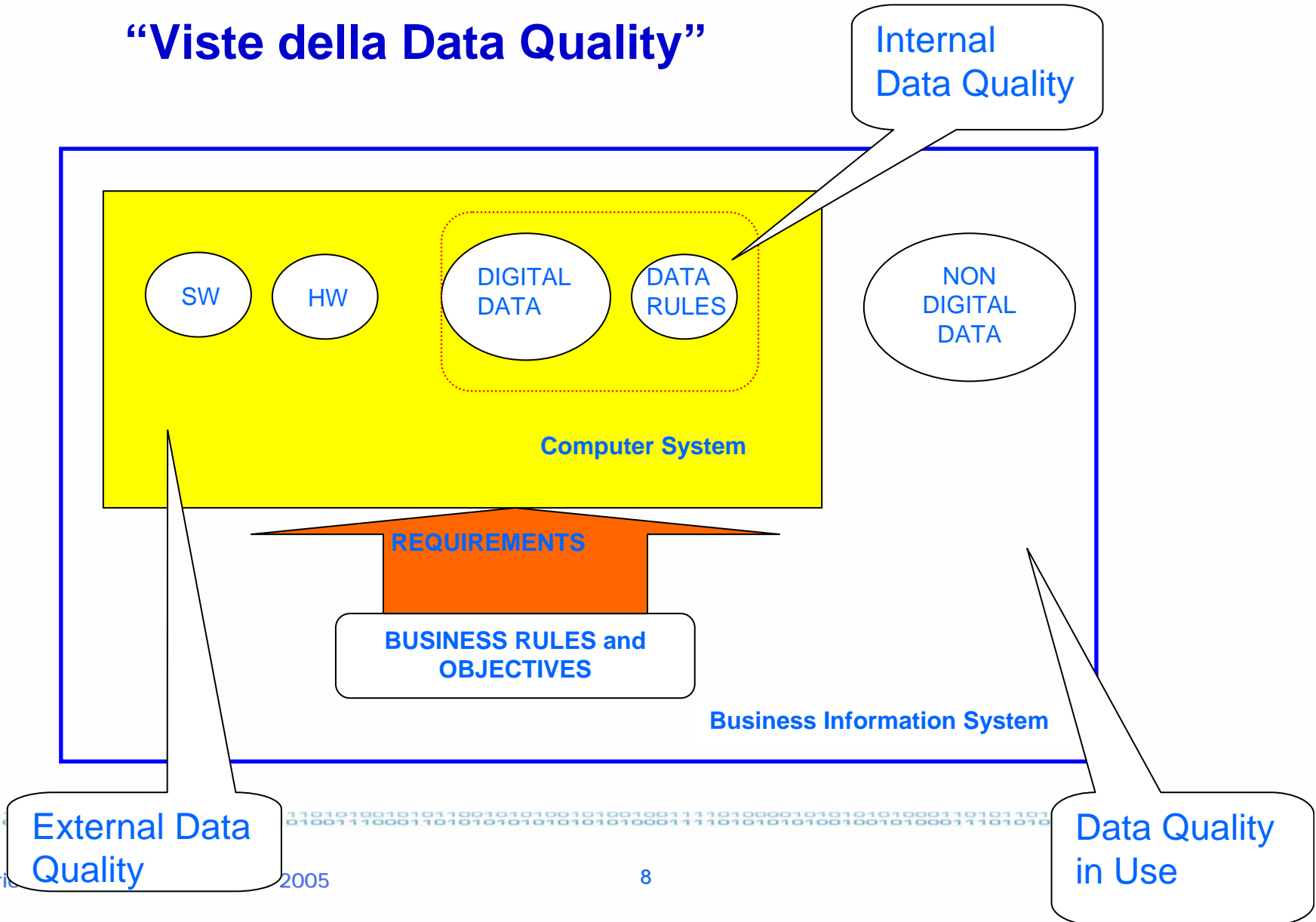
Le norme che fanno parte di questa “divisione” forniscono requisiti, raccomandazioni e linee guida per la valutazione del prodotto software, eseguita da valutatori, acquirenti o sviluppatori.

01010101111000011010101110101011101010010101100101010010100100111101000010101010100011010110100  
000111010111010101001010101010100011100011010101010101010001111010101010010010100011101010101

# ISO/IEC 25012 – Data Quality Model

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101010101101011010101001010101010100011100011010101010101010001111010101010010010100011101010101

## “Viste della Data Quality”



01010101  
00011101

11010100101011001010100101001001111010000101010101000110101101  
01001110001101010101010101010001111010101010010010100011101010



# Il modello ISO/IEC 25012

0101010111100001101010111100001101010111010101110101001010110010101001010010011110100001010101000110101101001  
000111010111010101011010111010101001010101010100111000110101010101010100001111010101010010010100011101010101

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)

010101011110000110101011101010111010100101011001010100101001001111010000101010100011010110100  
0001110101110101010010101010101001110001101010101010100001111010101010010010100011101010101





# Caratteristiche (2/4)

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. Questo standard si riferisce alla qualità dei dati, non del sistema: la caratteristica di ripristinabilità si riferisce pertanto ai singoli dati che è necessario siano ripristinabili. (ad es. attraverso meccanismi come: commit/synchpoint, rollback e backup-recovery).
	Conformità alle norme di affidabilità	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 Nota : 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 di usabilità	Deve essere possibile accedere ai dati con tempi di elaborazione e di throughput adeguati Nota: può essere riferito sia alle performances dell'HW o del SW sia alla qualità del disegno logico/fisico dei dati
Efficienza	Tempo di risposta	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
	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 di efficienza	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

# Caratteristiche (3/4)

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
000111010111010101010101010101010011100011010101010101010001111010101010010010100011101010101

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 alle norme di modificabilità	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 alle norme di portabilità	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la portabilità.

01010101111000011010101110101011101010010101100101010010100100111101000010101010100011010110100  
00011101011101010100101010101010011100011010101010101010001111010101010010010100011101010101

# Caratteristiche (4/4)

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101

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 di efficacia	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 di produttività	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la produttività.
Safety	Safety	La capacità dei dati di garantire un livello accettabile di rischio di danno alle persone, al business in uno specifico contesto. Esempio: i dati relativi al gruppo sanguigno di un individuo hanno rilievo per la safety
	Conformità alle norme di safety	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 di soddisfazione	La proprietà dei dati di essere conformi agli standard, alle convenzioni, norme o leggi che impongono prescrizioni circa la soddisfazione.

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101

# ISO/IEC 25012 – Data Quality Model

## “Viste” della Qualità dei Dati: qualità INTERNA

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

## 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, quantitiveness, convenience of access <sup>a</sup>
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)*

# Conclusione

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101010101101011101010101010100011100011010101010101010001111010101010010010100011101010101

In SOGEI sono state individuate tre categorie di 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

01010101111000011010101110101011101010010101100101010010100111101000010101010100011010110100  
000111010111010101001010101010100011100011010101010101010001111010101010010010100011101010101

# Conclusione

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
00011101011101

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)

01010101111000011010101110101011101010010101100101010010100100111101000010101010100011010110100  
000111010111010101001

# Bibliografia ...



- D. Natale, Qualità e Quantità nei Sistemi Software, FrancoAngeli, 1995.
- ISO/IEC 9126-1, Tecnologia dell'Informazione: Valutazione del prodotto software – Caratteristiche di qualità e linee guida per il loro uso, 2000.
- ISO/IEC 25012, Data Quality Model, 2005.
- D. Natale, Qualità, accessibilità e fruibilità delle informazioni, U&C Unificazione & Certificazione, Rivista UNI, Gennaio 2005.
- L. P. English, Improving Data Warehousing and Business Information Quality, Wiley, 1999
- D. Natale, M. Scannapieco, P. Angeletti, T. Catarci, G. Raiss, “Qualità dei dati e standard ISO/IEC 9126: Analisi critica ed esperienze nella Pubblica Amministrazione Italiana”, Workshop AIPA “Sistemi in rete nella Pubblica Amministrazione” with the participation of Sogei-Società Generale d’Informatica S.p.A. and the University of Rome “La Sapienza”, (in conjunction with The VLDB Very Large Data Base Conference 2001) Rome September 2001.Esempi di misurazioni
- Redman T.C.: Data Quality for the Information Age. Artech House, 1996
- Wang R.Y., Strong D.M.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, vol. 12, no. 4, 1996.
- ISO/IEC 19113 “Geographic Information – Quality principles” prepared by ISO/TC 211



# ... Bibliografia

010101011110000110101011110000110101011101010111010100101011001010100101001001111010000101010101000110101101001  
0001110101110101010110101110101010011100011010101010101010001111010101010010010100011101010101

- Azuma M.: Square, the Next Generation of the ISO/IEC 9126 and 14598 International Standard Series on Software Product Quality. In Proceedings of the 12th European Software Control and Metrics Conference, London, 2001
- Pipino, L., Lee, Y., Wang, R. "Data Quality Assessment" Communications of the ACM April 2002/Vol. 45, No. 4
- Olson, J. E. Data Quality: the accuracy dimension. Ed. Morgan Kaufmann Publishers. 2003
- ISO/IEC 2382-1 : Information technology – vocabulary – part 1 : fundamental terms
- ISO/IEC 9126: Information Technology- Software Product Evaluation- Quality Characteristics and Guidelines for their Use
- Huang, K.T., Lee, Y., Wang, R. Quality Information and Knowledge. Prentice Hall, Upper Saddle River, 1999
- Sarbanes-Oxley Act; Data Quality Act of 2001, Public Law 106-554
- ISO/IEC 11179 “Information Technology – Metadata registries (MDR)”

01010101111000011010101110101011101010010101100101010010100100111101000010101010100011010110100  
000111010111010101001010101010100111000110101010101010001111010101010010010100011101010101