



# ISO/IEC CD 25012 : DATA QUALITY MODEL

**Relatore:** Domenico Natale

Cnipa 10 Aprile 2006



## Sommario

- [Contesto del Sistema Informativo Fiscale](#)
- [Contesto generale](#)
- [Il progetto ISO : SQuaRE](#)
- [ISO-IEC CD 25012: definizioni](#)
- [La qualità interna, esterna ed in uso](#)
- [Conclusione](#)
- [Bibliografia](#)

Cnipa 10 Aprile 2006

2



## Contesto del Sistema Informativo Fiscale

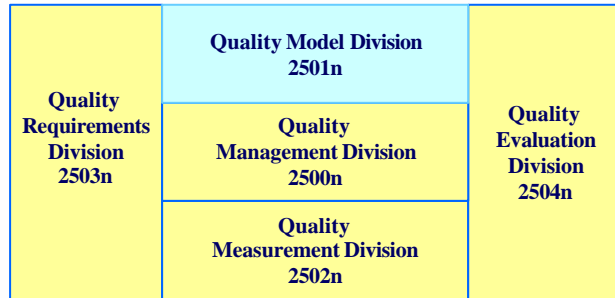
- Sogei, partner tecnologico del Ministero dell'economia e delle finanze, presta particolare attenzione ai temi della qualità dei dati e delle informazioni per la grande platea di utenza del Sistema Informativo Fiscale:
  - ✓ 50 sistemi informativi connessi
  - ✓ 1.200 applicazioni software integrate
  - ✓ 2.000 uffici dell'amministrazione collegati al sistema centrale
  - ✓ 8.000 comuni collegati
  - ✓ 140.000 intermediari utenti dei servizi telematici
  - ✓ 700.000 contribuenti utenti dei servizi telematici
  - ✓ 6 milioni di transazioni real-time giornaliere del sistema centrale
  - ✓ oltre 40 milioni di accessi annuali ai portali web del fisco
  - ✓ oltre 200 milioni di documenti trattati annualmente

## Contesto generale

- "Ogni generazione sarà più digitale di quella che l'ha preceduta" (Nicholas Negroponte, Essere digitale, 1995)
- Ogni sistema sarà sempre più integrato e interoperabile
- Il numero di banche dati sta crescendo più delle applicazioni sw
- L'esigenza di elevato livello di garanzia di qualità del dato è in aumento
- In ambito ISO la qualità dei dati è oggetto di sempre maggiore attenzione (11179 sui metadati e 19113 sul GIS)
- Gli utenti internet a livello mondiale (e le banche dati scambiate) crescono in continuazione:
  - ✓ 2000: 195 milioni
  - ✓ 2002: 495 milioni
  - ✓ 2005: 965 milioni

## Il progetto ISO : SQuaRE

la Qualità dei Dati è inserita nel progetto ISO **SQuaRE** *Software Product Quality Requirements and Evaluation*



**SOGEI** partecipa come editor al gruppo di lavoro ISO JTC1/SC7/WG6 (Evaluation and metrics) che sta redigendo la norma sulla Qualità dei Dati (ISO/IEC 25012) nell'ambito della Quality Model Division

## 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.
- 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.

## Il Progetto ISO SQuaRE

- 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.

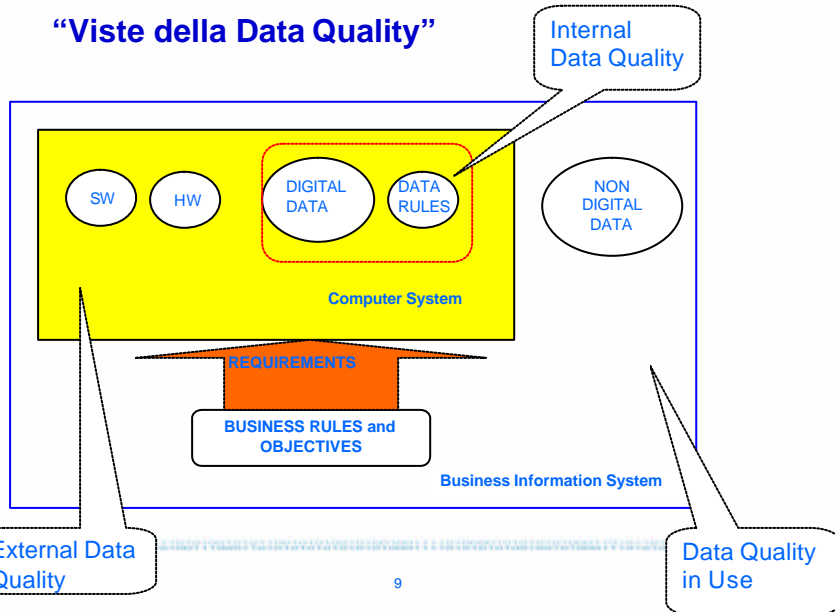
Le norme che fanno parte di questa “divisione” forniscono requisiti, raccomandazioni e linee guida per la valutazione del prodotto software, eseguita da valutatori, acquirenti o sviluppatori.

## ISO-IEC CD 25012: definizioni

- Il dato è una rappresentazione reinterpretabile di una informazione
  - ✓ I dati possono essere elaborati sia da esseri umani che da macchine (2382)
  - ✓ Ogni dato contiene un grado di informazione
- Per informazione si intende una parte della conoscenza di oggetti come ad esempio fatti, eventi, processi, idee, concetti, che hanno un particolare significato in un determinato contesto (2382)
- Per qualità dei dati si intende la capacità dei dati di soddisfare esigenze sia esplicite che implicite se usati sotto determinati condizioni
- Per metadato si intende un dato che descrive un altro dato (11179)
- Qualità interna: valori, formati, definizioni, regole, relazioni
- Qualità esterna: ambiente tecnologico, caratteristiche hw e sw
- Qualità in uso: contesto d'uso dell'utente (efficacia, produttività, safety, soddisfazione)

## ISO/IEC 25012 – Data Quality Model

### “Viste della Data Quality”



Cnipa 10

9

## Il modello ISO/IEC 25012

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

# Il modello ISO/IEC 25012

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

## Caratteristiche (1/4)

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"
	Accuratezza (semantica e sintattica)	E' il grado di conformità di un valore acquisito, misurato e calcolato, al suo attuale valore. E' caratterizzata principalmente da due aspetti 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 quando il nome Maria è riportato come Marja 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 quando il nome Maria è riportato come Francesca. Francesca è sintatticamente accurato, perché è compreso nel dominio di riferimento, ma è un nome diverso.
	Interoperabilità	E' possibile accedere a scambiare i dati è 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 (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, rollbacke backup-recovery).
	Conformità alle norme di affidabilità	La proprietà dei dati di essere conformi agli standard, alle convenzioni, no rme 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)

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, no rme 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, no rme o leggi che impongono prescrizioni circa la portabilità.

## Caratteristiche (4/4)

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.

## La qualità interna, esterna ed in uso

### “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
- ✓ Ha 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

## La qualità interna, esterna ed in uso

### “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.
- ✓ Esempio: 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”.

## La qualità interna, esterna ed in uso

### “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*
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

Sono state individuate tre categorie di strumenti:

Strumenti di analisi della DQ e di indagine

- 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

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

Strumenti di gestione

- 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. con XML) e accessibile (L. Stanca)
- ✓ Documenti, studi e analisi, interfaccia web
- ✓ Record, banche dati integrate, rete internet
- ✓ Metadati, cataloghi e dizionari
- ✓ Applicazioni software specifiche (es. protocollo)
- ✓ Sistemi informatici
- ✓ Sistemi informativi
- ✓ Conoscenza digitale (KM) garantita per l'oggi (business continuity e backup) e per il domani (conservazione permanente)

## 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

- 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)"