

LA QUALITÀ DEI DATI E L'INFORMAZIONE STATISTICA

D. NATALE*, M.C. PAOLETTI**, A. SIMONETTA***

SOMMARIO

1. Introduzione. - 2. La qualità nell'informazione statistica. - 2.1 Il controllo dei dati. - 3. La serie di standard SQuARE. - 3.1 Architettura della serie ISO/IEC 25000. - 4. Conclusioni.

1. Introduzione

L'evoluzione dell'ICT richiede in modo sempre più pressante l'incremento delle performance, fenomeno che è accompagnato dalla graduale diminuzione dei costi di hardware e software. La rapida crescita dell'uso dei servizi internet accentua l'importanza dell'interazione dell'utente con il sistema (usabilità), dell'interazione tra sistemi (interoperabilità), dell'integrità con sistemi aperti (sicurezza) e del governo delle applicazioni *mission critical* (affidabilità). In questo scenario la qualità del software e dei dati è divenuta cruciale.

Tante sono le motivazioni che spingono ancora il mondo ISO, dopo oltre trenta anni di ingegneria del software, a far evolvere gli standard attuali verso nuove prospettive che tengano conto di modelli di qualità del prodotto.

Il progresso tecnologico non è sempre sinonimo di semplificazione: l'aumento della complessità dei sistemi informatici è talvolta causata dalla diminuzione della qualità del software e dell'informazione o, anche, dal fatto che tale requisito è difficile da controllare.

La qualità nei dati è una componente principale di quella dell'informazione e molti processi della vita sociale dipendono proprio da essa. Infatti, la bassa qualità sui dati contribuisce a fornire informazioni non accurate, a risultati non utilizzabili e a rendere gli utenti insoddisfatti. Inoltre si sperimenta di continuo che

* Presidente Commissione ISO/IEC JTC1/SC7 UNINFO.

** Consulenza Statistico Attuariale - Direzione Generale INAIL.

*** Consulenza per l'Innovazione Tecnologica - Direzione Generale INAIL.

non servono dati di qualità se il software che li gestisce rimane farraginoso e complesso, come è inutile disporre di un software semplice ed usabile se poi i dati resi fruibili sono incoerenti.

La quantità di applicazioni software e dei dati gestiti dai sistemi informatici sta aumentando in tutto il mondo. Secondo le stime dell'azienda CISCO nella fine del 2015 la quantità di traffico IP su internet raggiungerà lo Zettabyte (che corrisponde ad un trilione di Gigabyte) con uno scambio dati per anno di circa 80,5 Exabyte per mese [18], stime meno prudenti [19] prevedono che nel 2020 l'universo digitale raggiungerà quota 35 Zettabyte (44 volte il valore del 2009). Questa impressionante crescita di informazioni richiederà l'utilizzo di modelli standardizzati che siano di ausilio al governo della complessità.

Inizialmente l'ingegneria del software si è occupata di misurare la complessità delle applicazioni software attraverso l'analisi statica del codice (numero delle LOC¹, commenti, istruzioni decisionali, strutture "one in/out", modularità) [2,4]. Fin dal 1990 si è data attenzione alla complessità delle interfacce grafiche e alla navigazione uomo-macchina (*Human Computer Interaction*, di seguito HCI) [3]. Questo approccio ha fatto nascere un nuovo modello di analisi che valutasse le reali funzioni consegnate all'utente finale indipendentemente dal modo attraverso il quale venissero realizzate (Punti Funzione o Function Point).

Attraverso la tecnica dei Punti Funzione, diffusamente utilizzata anche nel campo delle acquisizioni della PA, è possibile stimare la dimensione e, conseguentemente, il costo delle applicazioni software da realizzare.

Studi più recenti tendono a completare l'analisi della complessità considerando anche il contenuto dell'informazione rappresentata nelle applicazioni software, il valore e significato del dato stesso [15] ma anche l'ergonomia e l'usabilità dell'interfaccia [17,20]. La complessità nel campo HCI non è solo dovuta alle difficoltà incontrate durante la navigazione e l'esecuzione di compiti, ma anche al fatto di incontrare errori nei dati forniti. Si sottolinea questo aspetto perché dati e metadati non corretti appaiono frequentemente nelle applicazioni web, nelle interfacce e nei servizi IT resi al cittadino. Spesso le cause sono banali e legate al disallineamento tra archivi differenti oppure tra i servizi IT ed i processi aziendali interni, ma sono il sintomo dell'assenza di controlli specifici.

Nei processi di produzione del software è fondamentale includere una fase di test specifica per la verifica della qualità sui dati, controllo che si concretizza nella misura dell'accuratezza, della completezza, della coerenza, dell'accessibilità e della comprensibilità così come definito nella sezione *La serie di standard SQuaRE*. Questi sono aspetti del contenuto dell'informazione e nulla hanno a che fare con gli artefatti software sviluppati per visualizzare l'informazione, seppur anch'essi importanti, da valutare durante la fase di progettazione.

1 LOC è acronimo di Line of Code cioè la metrica utilizzata per misurare la dimensione di un'applicazione software in termini delle linee di codice sorgente che la compongono.

Come afferma Norman [1] la conoscenza è nella mente umana, mentre l'informazione è fuori di essa: nei testi scritti e negli elementi culturali. Per questa ragione è importante verificare che l'informazione sia il più possibile conforme a caratteristiche attese di qualità, riducendo confusioni o stress legati a difficoltà d'interpretazione nei processi di elaborazione cognitiva. Ciò significa che deve esistere un'adeguata coerenza tra l'informazione percepita dall'utente e quella contenuta nelle banche dati. Quindi l'utente deve avere una visione integrata della realtà indipendentemente dal fatto che le informazioni possono essere estratte da fonti diverse.

Informazioni errate, incoerenti, o anche solo approssimate, sono spesso presenti in vari settori della società dell'informazione: segnali stradali "fuori standard", cartelli con simbologia contraddittoria in aree comuni (parchi, aeroporti, ospedali), indicazioni sui mezzi pubblici carenti o con formati differenti,...

Questo problema nella rappresentazione delle informazioni e, conseguentemente, della scarsa qualità nei dati, è ancora più evidente se si mettono a confronto contesti differenti: Paesi (o città) di diverse aree geografiche pur condividendo aspetti legati alla vita sociale, giuridica ed economica, talvolta utilizzano differenti sistemi di codifica.

Per migliorare i processi che sottendono la qualità del software è necessario utilizzare modelli di rappresentazione che si avvicinino agli stereotipi degli utenti finali, minimizzando la distanza concettuale tra percezione e modello mentale. L'adozione di un modello di qualità che definisce le caratteristiche da considerare permette di anticipare i possibili errori di interpretazione, evidenziando cause e costi della carenza di qualità.

La misura della qualità del prodotto software e dei dati è una parte essenziale di questa analisi ed il percorso logico da seguire per orientarsi secondo questa nuova prospettiva è:

- muovere la cultura da standard di processo (certificabile convenzionalmente da terzi) a standard di prodotto, oggetto valutabile dall'utente in termini reali e di consumo;
- dare importanza ai requisiti iniziali e alle valutazioni dell'utente sui risultati conseguiti;
- applicare modelli standard di qualità;
- spostare l'attenzione produttiva dal punto di vista del progettista software al punto di vista dell'utente finale che dovrebbe ricevere servizi informatici e dati coerenti di qualità;
- misurare i diversi livelli di qualità raggiunta nelle varie fasi del processo di gestione (test, esercizio, manutenzione);
- indagare sulle cause che determinano la scarsa qualità identificando i fattori di rischio;
- migliorare la qualità nei processi interni ed esterni;

- assicurare la qualità inter-amministrativa e delle varie sorgenti da cui si alimentano i flussi di scambio dati;
- utilizzare standard di interscambio dei dati;
- concentrarsi dapprima su pochi dati di maggior rilievo per l'utente finale e per il contesto di utilizzo (master data), solo successivamente arricchire la base di conoscenza con quelli meno significativi;
- fornire dizionari semplici per l'eliminazione delle ambiguità semantiche.

I modelli che saranno descritti sono il risultato concreto di iniziative di ricerca, attività nel mondo produttivo e di “*best practice*” italiane ed internazionali.

2. La qualità nell'informazione statistica

Con il recente fenomeno dell'Open Data [16], le Pubbliche Amministrazioni sono sempre più chiamate ad esporre, online ed in formato aperto, il proprio patrimonio informativo (nel rispetto delle norme riguardo alla riservatezza dei dati).

La diffusione di internet rappresenta una grande facilitazione rispetto ai tempi passati e gli utilizzatori stanno aumentando ogni giorno. Nascono anche nuovi fenomeni come l'Open Journalism che consentirà ai giornalisti di pubblicare articoli su dati ufficiali ed estratti dagli archivi aperti delle organizzazioni pubbliche e private.

In questo ambito le PA forniscono anche elaborazioni statistiche, a vari livelli di astrazione e granularità, ai fini della divulgazione di informazioni di sintesi rispetto ai dati che la PA stessa gestisce. Per questa ragione è necessario che i processi di produzione statistica seguano, in conformità con gli standard europei, determinate regole e principi. Proprio per promuovere il miglioramento della qualità delle statistiche prodotte all'interno del Sistema Statistico Nazionale, è stato recentemente promulgato il Codice italiano della Statistica ufficiale (G.U. n. 240/2010). Inoltre, nel giugno 2011 l'Istituto Nazionale di Statistica ha definito le *Linee guida per la qualità dei processi statistici* con lo scopo di divulgare i principi base cui attenersi nella progettazione, nell'esecuzione e nel controllo di un'indagine statistica adottando la definizione di qualità, raccomandata dal LEG on Quality e da Eurostat, secondo cui la qualità è definita come “*il complesso delle caratteristiche di un prodotto o di un servizio che gli conferiscono la capacità di soddisfare i bisogni impliciti o espressi*” [5,6]. Anche in questo ambito esistono due possibili approcci a cui fare riferimento: la qualità di processo e di prodotto finale.

La qualità di processo si ottiene agendo sulle varie fasi del ciclo di produzione statistica definendo l'obiettivo e le regole (linee guida da seguire) per conseguire quell'obiettivo. Ogni indagine dovrà quindi prevedere la definizione dell'obiettivo, del disegno progettuale, dell'archivio delle unità di rilevazione, del trattamento dei dati (acquisizione, archiviazione e diffusione) e della documentazione finale.

La qualità sui prodotti si attua attraverso la definizione di requisiti di qualità che possono essere poi misurati sui prodotti finali con lo scopo di conoscerne il livello di qualità raggiunto.

Dal punto di vista strettamente statistico i prodotti di un processo possono essere di tre tipologie: *microdati*, *macrodati* e *metadati*. Sulla scia delle ultime disposizioni legislative la pubblica amministrazione è tenuta a ridurre al minimo le informazioni da richiedere ai cittadini attraverso l'integrazione delle informazioni tra più archivi.

Questo si realizza a livello di microdati con l'abbinamento di un codice esatto o attraverso record linkage deterministico; a livello meta, attraverso l'abbinamento di unità "fittizie" ottenute per aggregazioni di dati provenienti da archivi diversi; a livello macro, attraverso diverse elaborazioni di dati sullo stesso fenomeno. L'integrazione a cui si fa riferimento è sia a livello di "contenuto", ottenuta da informazioni provenienti da sorgenti dati diverse, sia di "processo", attraverso il riuso di medesime procedure o dei risultati provenienti da diversi contesti produttivi statistici. Affinché si possa realizzare l'integrazione di contenuti è necessario aver formalizzato e condiviso requisiti essenziali quali le definizioni statistiche, le classificazioni e le matrici dei dati. La qualità dell'informazione statistica è quindi, strettamente legata a garanzie di progettazione (rilevanza, accuratezza statistica, tempestività, accessibilità, comparabilità e coerenza) e garanzia di tolleranza (precisione campionaria, riduzione dell'errore non campionario,...). Secondo le definizioni ufficiali di Eurostat, tradotte poi nelle linee guida dell'Istituto Nazionale di Statistica, le principali componenti della qualità sono:

- la rilevanza che rappresenta il grado in cui l'informazione soddisfa le esigenze attuali e potenziali degli utenti;
- l'accuratezza che rappresenta il grado di vicinanza tra le stime e i reali valori;
- la tempestività dei risultati che è definita come il periodo di tempo che intercorre tra il fenomeno e il momento in cui gli stessi sono resi disponibili; mentre la puntualità rappresenta il periodo di tempo che intercorre tra la data di rilascio dei dati e quella di rilascio programmata;
- la coerenza che è riferita alle stime e misura quanti i processi statistici che le hanno prodotte utilizzano metodi armonizzati e concetti comuni (classificazioni, definizioni e popolazioni obiettivo); la comparabilità, è un caso particolare della coerenza, e indica la possibilità di eseguire confronti attendibili tra statistiche di domini (spaziali e temporali) diversi;
- l'accessibilità che è definita come la facilità per gli utenti finali di accedere ai dati (come e dove richiederli, politica di diffusione e prezzi, formati di memorizzazione,...); mentre la chiarezza rappresenta la semplicità per gli utenti di comprendere i dati.

Per le definizioni ufficiali promosse da Eurostat si rimanda alla lettura della documentazione di riferimento [12].

Tuttavia, misurare la qualità di un prodotto statistico non è un compito semplice,

alcune caratteristiche sono di natura quantitativa (tempestività, accuratezza e comparabilità), altre di natura qualitativa (rilevanza, accessibilità e chiarezza)[6,7]. Inoltre, anche la misurazione di una caratteristica quantitativa come l'accuratezza può presentare delle insidie, infatti, essa dipende in modo inversamente proporzionale dal numero di errori che possono emergere durante il processo di elaborazione statistica. In generale, esistono due diverse tipologie di errore: campionario e non campionario. Il primo è legato alla dimensione del campione, al disegno di campionamento ed allo stimatore utilizzato per derivare le stime delle quantità di interesse. Gli errori non campionari generano la distorsione delle stime rispetto ai valori reali, tra i tanti si citano quelli di copertura, di mancata risposta, di misurazione, materiali, di progettazione e, infine, gli errori causati dall'influenza dei soggetti intervistati durante l'espressione del giudizio. Gli errori di copertura nascono da errori nelle liste delle unità da osservare e possono introdurre distorsione nei casi di sottocopertura (duplicazione dati, errori nei dati di contatto,...) oppure aumento della variabilità delle stime nei casi di sovracopertura (a causa della necessaria riduzione selettiva del campione). Gli errori di mancata risposta sono quelli che non consentono di rilevare l'unità di osservazione. Gli errori di misurazione si concretizzano quando, al termine di una elaborazione, il valore di un dato si discosta dal quello reale. Gli errori materiali possono accadere durante le fasi di codifica/decodifica delle informazioni oppure durante la fase di trascrizione dei valori. L'errore totale d'indagine è quello che racchiude in se tutti i possibili errori dell'indagine ed è, senza ombra di dubbio, quello più complesso da stimare. Tale complessità dipende dal fatto che i vari errori di un'indagine statistica non sono entità indipendenti ed è difficile trovare un modello statistico esaustivo che ne descriva le relazioni.

2.1 Il controllo dei dati

Il controllo dei dati permette di individuare e correggere gli errori riscontrati nelle unità statistiche durante le varie fasi di produzione statistica. I metodi, gli strumenti e le procedure di controllo e correzione devono essere scelti tenendo conto delle esperienze disponibili su dati simili, delle linee guida, degli standard (o raccomandazioni) sviluppati a livello nazionale o internazionale.

Per identificare gli errori esistono due differenti tipologie di controlli: qualitativi e quantitativi. I primi rilevano gli scostamenti dei prodotti statistici rispetto alle fonti certe. Invece, i controlli qualitativi cercano relazioni inaspettate tra le variabili o sui dati (valori mancanti, anomali o fuori intervallo) al fine di valutare lo stato dell'archivio rispetto alle informazioni contenute.

Secondo le linee guida dell'Istat, i dati mancanti (*missing value*) devono essere riconoscibili rispetto ai valori non dovuti e, nel caso di variabili quantitative, anche rispetto agli zeri strutturali.

Dall'analisi degli indicatori relativi alle procedure di controllo devono essere definite le regole atte all'individuazione degli errori di natura sistematica che devono essere identificati e corretti prima degli errori casuali e dell'editing selettivo.

Per il riconoscimento dei valori anomali devono essere utilizzati metodi che individuino le relazioni tra le variabili come le analisi univariate. Infine, per gli errori casuali dovrebbe essere utilizzato un software che sfrutti una metodologia consolidata come quella basata sul "principio di minimo cambiamento".

In generale, tutte le procedure di controllo e correzione dovrebbero essere automatizzate attraverso software flessibili, per quanto concerne le possibili modifiche o integrazioni, che siano la traduzione di approcci e metodi consolidati per quel particolare ambito statistico. Alla luce di quanto esposto risulta fondamentale preservare i dati originali ai fini del controllo sulle stime effettuate.

3. La serie di standard SQaRE

Il passaggio dalla qualità di processo alla qualità di prodotto è una novità in campo ISO che si concretizza con la serie di standard SQaRE - acronimo di *Software product Quality Requirements and Evaluation* (ISO2/IEC3 25000)[8] - risultato di un'ampia attività internazionale di esperti del settore.

Le norme contenute nella serie SQaRE si basano su modelli di qualità fondamentali e comprendono aspetti riguardanti i requisiti, la valutazione, la gestione e la misurazione.



Fig. 1: Elementi di "SQaRE" ISO/IEC 25000.

Tali standard si prefiggono di focalizzare l'attenzione non solo verso i requisiti, ma anche verso le fasi finali della valutazione e misurazione della qualità, partendo da modelli specifici di qualità che non riguardano solo aspetti di qualità interna ed esterna, ma soprattutto di quella percepita dall'utente finale (qualità in uso).

2 ISO è l'International Organization for Standardization (www.iso.org)

3 IEC è l'International Electrotechnical Commission (www.iec.ch)

In Italia l'attività di normazione ISO è seguita dall'UNI⁴, che per la parte ICT si avvale dell'Ente Federato UNINFO⁵, ove sono accreditati rappresentanti dell'Istituto. Grazie a tali National Body anche gli esperti italiani possono collaborare con i colleghi di altri Paesi e condividere la realizzazione e l'aggiornamento delle norme tecniche.

Tutte le definizioni ed i termini relativi alla serie SQaRE utilizzati nel presente documento sono il frutto della traduzione delle norme ISO, pertanto devono essere considerati come indicativi. Il riferimento ufficiale è l'insieme degli standard ISO pubblicati nella versione originale inglese [10,14].

Nell'ambito della trattazione della serie SQaRE verrà data particolare enfasi agli standard *System and software quality models* (ISO/IEC 25010) e *Data Quality Model* (ISO/IEC 25012) a ragione del cospicuo contributo fornito dagli esperti italiani.

3.1 Architettura della serie ISO/IEC 25000

Lo sviluppo degli standard ISO/IEC 25000 è iniziata all'inizio degli anni 2000 e oggi si trova in diversi stadi del processo di realizzazione e di pubblicazione. Lo schema di fig. 2 fornisce un quadro di sintesi dell'architettura e degli standard che la compongono (M. Azuma, Waseda University, Tokyo), sono riportate le norme già pubblicate (fondo verde) e quelle in corso di realizzazione (fondo celeste).

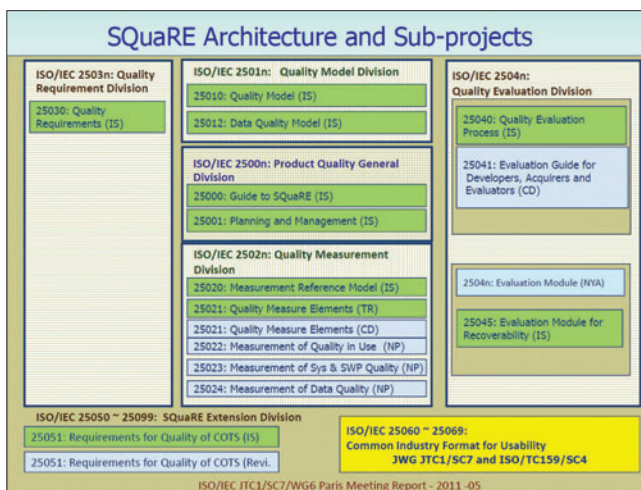


Fig. 2: Architettura e progetti "SQaRE" ISO/IEC 25000.

- 4 UNI è l'Ente Nazionale Italiano di Unificazione (www.uni.com)
- 5 UNINFO è l'Ente di formazione federato all'UNI che promuove e partecipa allo sviluppo della normativa nel settore delle tecniche informatiche (www.uninfo.polito.it)

La qualità del software, descritta nella *Guide to SQuaRE* del 2005 (ISO/IEC 25000), è considerata come la capacità del prodotto, usato in condizioni specifiche, di soddisfare le esigenze implicite ed esplicite dell'utente finale. Dove per esigenze dell'utente si intendono le aspettative che egli ha durante l'utilizzo del prodotto. Questa norma è molto importante poiché fornisce un quadro d'insieme del progetto SQuaRE e una guida per l'utente nell'utilizzo degli standard della serie. Tra i maggiori contenuti si evidenziano quelli tesi a fornire relazioni con i precedenti documenti in ambito ISO (ISO/IEC 9126 e 14598), le definizioni terminologiche e l'architettura generale.

Nello standard *Planning and management* del 2007 (ISO/IEC 25001) sono descritti i dettagli orientativi per la pianificazione, per la gestione dei requisiti e per la valutazione. Il linea con l'ISO/IEC 15288 (definizione e analisi dei requisiti di qualità), le organizzazioni che vogliono assicurare la qualità sul prodotto software e la relativa valutazione di conformità, hanno nella norma una chiara evidenza dei requisiti da identificare. Lo standard *System and software quality models* del 2011 (ISO/IEC 25010) è adottabile all'interno di qualsiasi metodologia che si prefigga di determinare una produzione del software con risultati di alta qualità. Secondo l'impostazione della norma la qualità è suddivisa in tre ambiti:

- *qualità interna* verificabile con ispezioni o con strumenti di valutazione delle proprietà statiche del codice sorgente;
- *qualità esterna* verificabile con test dinamici sulle applicazioni in ambienti simulati;
- *qualità in uso* verificabile in ambienti reali, o simulati, con la partecipazione di utenti che evidenziano le difficoltà o la semplicità nell'uso dell'interfaccia uomo-macchina.

La tabella a pagina seguente sintetizza il modello della qualità interna-esterna del prodotto software e del sistema che lo ospita.

Tabella 1

Caratteristiche e sotto-caratteristiche del modello di “qualità interna/esterna” ISO/IEC 25010.

<p>Idoneità funzionale <i>Completezza</i> <i>Correttezza</i> <i>Appropriatezza</i></p>	<p>Affidabilità <i>Maturità</i> <i>Disponibilità</i> <i>Tolleranza ai guasti</i> <i>Recuperabilità</i></p>
<p>Efficienza <i>Temporale</i> <i>Utilizzazione risorse</i> <i>Capacità</i></p>	<p>Sicurezza <i>Riservatezza</i> <i>Integrità</i> <i>Non ripudio</i> <i>Responsabilità</i> <i>Autenticità</i></p>
<p>Compatibilità <i>Coesistenza</i> <i>Interoperabilità</i></p>	<p>Manutenibilità <i>Modularità</i> <i>Riusabilità</i> <i>Facilità di analisi</i> <i>Modificabilità</i> <i>Testabilità</i></p>
<p>Usabilità <i>Riconoscibilità appropriata</i> <i>Facilità di apprendimento</i> <i>Operabilità</i> <i>Protezione da errori utente</i> <i>Estetica dell'interfaccia utente</i> <i>Accessibilità</i></p>	<p>Portabilità <i>Adattabilità</i> <i>Facilità di installazione</i> <i>Sostituibilità</i></p>

A monte del processo di produzione, occorre definire l'importanza delle singole caratteristiche in funzione del contesto d'uso e delle priorità assegnate. La tabella seguente fornisce un quadro del modello della qualità in uso, in questo caso l'importanza delle caratteristiche è definita in base all'opinione generale degli utenti.

Tabella 2

Caratteristiche e sotto-caratteristiche del modello di “qualità in uso” ISO 25010.

<p>Efficacia <i>Efficacia</i></p>	<p>Assenza e attenuazione rischi <i>Rischio economico</i> <i>Rischio per le persone</i> <i>Rischio danno ambientale</i></p>
<p>Efficienza <i>Efficienza</i></p>	
<p>Soddisfazione <i>Utilità</i> <i>Fiducia</i> <i>Piacere</i> <i>Comodità</i></p>	<p>Coperatura del contesto <i>Completezza del contesto</i> <i>Flessibilità</i></p>

La qualità in uso di un software o di un sistema misura l'impatto che il prodotto ha sull'utente e dipende dalla qualità del software, dall'hardware, dall'ambiente operativo e dai compiti dell'utente e suo contesto sociale.

Il "Data quality model" nasce nel 2008 (ISO/IEC 25012) e classifica gli attributi di qualità in quindici caratteristiche secondo due differenti punti di vista:

1. *Inerente ai dati*: la qualità del dato si riferisce al valore del dato stesso, a prescindere dal supporto di rappresentazione;
2. *Dipendente dal sistema*: la qualità del dato è ricercata e preservata all'interno di un sistema informatico; essa dipende dal dominio tecnologico che gestisce i dati (hardware, dispositivi di memorizzazione, software, ecc.).

Alcune di queste caratteristiche appartengono ad uno specifico punto di vista mentre altre hanno impatto su entrambi gli aspetti. Generalmente gli aspetti inerenti ai dati sono di responsabilità degli esperti del dominio, mentre quelli dipendenti dal sistema di responsabilità dei tecnici.

La fig. 3 fornisce un quadro di sintesi delle quindici caratteristiche organizzate rispetto al punto di vista ed al dominio di appartenenza (fatti ed artefatti) [11,13]. Questa figura è una elaborazione concettuale che non appartiene alla versione originale ISO che invece riporta il semplice elenco delle caratteristiche.

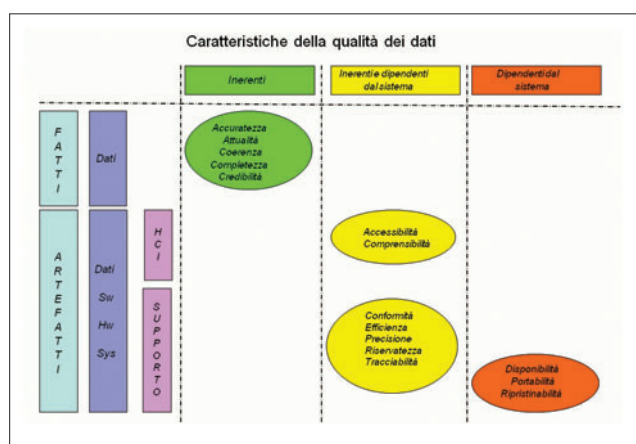


Fig. 3: Modello concettuale di Data quality.

Le caratteristiche relative al punto di vista inerente ai dati sono connesse ai fatti della realtà, quelle dipendenti dal sistema sono connesse ad artefatti tecnologici che ospitano i dati (supporti, dispositivi e ambienti), le altre sono connesse ad entrambi gli aspetti. Inoltre, alcune di esse sono in relazione con argomenti attua-

li dell'ICT: l'accessibilità e l'usabilità sono aspetti legati alla disciplina della HCI; la riservatezza e la tracciabilità con il Codice della Privacy; la disponibilità, la portabilità e la ripristinabilità con il Disaster recovery, la Business continuity ed il Cloud computing.

L'importanza di controllare la qualità dei dati è un fattore cruciale per migliorare la qualità dei processi: i migliori risultati si ottengono con misurazioni eseguite in automatico ed in tempo reale, utilizzando basi di dati per la memorizzazione dei trend e grafici multidimensionali per la rappresentazione globale del fenomeno (es. diagrammi radar). Seguendo il concetto di "Total quality", dove l'utente che crea un dato deve essere in grado di seguirlo - dal livello di dettaglio a quello di estrema sintesi - nella filiera produttiva, l'adozione della filosofia open data appare molto utile anche per garantire la trasparenza nel trattamento dei dati. Per garantire livelli di qualità elevati è necessario controllare significati e contenuti dei dati, ma altresì disporre di una infrastruttura tecnologica adeguata.

Il modello di riferimento per la misurazione delle caratteristiche di qualità in SQuaRE è descritto nello standard *Measurement reference model and guide* del 2007 (ISO/IEC 25020). La norma contiene i criteri che possono essere adottati per identificare le misure di qualità (QM=Quality Measure) e gli elementi oggetto di misurazione (QME=Quality Measure Element).

I requisiti e le raccomandazioni per le specifiche della qualità del software sono descritti nello standard *Quality requirements* pubblicato nel 2007 (ISO/IEC 25030). Lo standard, che si applica indistintamente ad acquirenti e fornitori, si concentra sui requisiti di qualità del software, tenendo conto della prospettiva di sistema: il software è normalmente sviluppato come parte di un mondo più ampio. Secondo questa impostazione i requisiti di qualità sono necessari per le fasi di specifica (inclusi gli accordi contrattuali), pianificazione (incluse le analisi di fattibilità), lo sviluppo (inclusa la tempestiva identificazione dei problemi che potrebbero verificarsi) e la valutazione (inclusi gli obiettivi di assessment e la certificazione). Infine, lo standard "Quality requirements" del 2011 (ISO/IEC 25040) tratta i requisiti e raccomandazioni per la valutazione del software. Il processo di misurazione può essere effettuato per valutare la qualità preliminare del software da sviluppare (ISO/IEC 14598-3), durante il processo produttivo (ISO/IEC 14598-4) oppure al termine della fase di produzione (ISO/IEC 14598-5).

La ISO/IEC 25040 nasce originariamente per i responsabili della fase di valutazione ma risulta fondamentale anche per altre figure professionali come gli sviluppatori software, gli acquirenti e gli auditor esterni.

4. Conclusioni

Esaminando approfonditamente i nuovi standard si possono individuare le caratteristiche che hanno accompagnato le problematiche dell'evoluzione dell'infor-

matica fin dal suo nascere. Inizialmente dai sistemi centrali proprietari degli anni '60-'70, ai PC degli anni '80, alle reti degli anni '90, ai sistemi sempre più evoluti dei tempi attuali (fissi e mobili) incentrati sull'utente finale, alle problematiche di questi anni relative alla qualità dei dati per il Cloud computing e per l'Open data [16], all'integrazione dei sistemi informativi [9].

Con la serie SQuaRE vengono enunciate esplicitamente le relazioni tra la qualità del software e la qualità dei dati, come già evidenziato da qualche anno da vari ricercatori italiani e per troppo tempo lasciate in termini impliciti nella storia del software engineering.

L'istituto partecipa attivamente ai vari *Working Group* ISO, UNINFO e UNI, con professionisti ed esperti che hanno assicurato la compatibilità degli standard internazionali con gli orientamenti e le migliori prassi della Pubblica Amministrazione e dell'Università, seguendo le linee nazionali di eGovernment e codici come il CAD e la Privacy.

Attualmente si sottolinea l'attenzione sulla progettazione in ambito ISO di nuovi Work Item: il 25022 relativo alle metriche di qualità in uso [21], il 25023 relativo alle metriche utili al controllo della qualità del software [22], il 25024 relativo alla misurazione della qualità dei dati [23]. Tutti i Work Item sono orientati a favorire la messa in pratica, nell'ambito reale produttivo, dei concetti fin ora esposti.

Tra le prospettive future il progetto SQuaRE si propone di agevolare l'integrazione degli strumenti CASE di gestione e lo sviluppo di misure basate su campi esperienziali avanzati. Molta attenzione sarà data alla coerenza del proliferare degli standard e alla coerenza con le precedenti norme, avendo come obiettivo prevalente quello di assicurare la qualità del prodotto e dei servizi, velocizzando anche il processo di standardizzazione.

L'eliminazione della carta nelle attività d'ufficio attraverso l'utilizzo di computer, della manualità di compilazione dei modelli, della complessità dei servizi burocratici può avvenire solo concertando un'azione comune della PA che semplifichi i regolamenti, il software, le banche dati e crei feedback più rapidi e lineari tra cittadino e amministrazioni, entrambi alla ricerca della coerenza informativa e dei dati reali.

RIASSUNTO

L'evoluzione dell'ICT presta attenzione non solo alla tradizionale qualità dei processi, ma anche alla qualità dei prodotti e quindi dei servizi, ponendo le esigenze dell'utente come primarie. Per questo motivo, al fine di verificare l'esattezza dell'informazione, sono stati introdotti opportuni test di controllo e dei modelli standard di qualità da seguire. Tali modelli si basano essenzialmente su aspetti relativi ai requisiti, alla valutazione, alla gestione e alla misurazione della

qualità. In tale ottica è stata prodotta la serie degli standard SQuaRE, risultato di un'ampia attività internazionale cui l'Istituto ha partecipato. In Italia l'attività di normazione è seguita dall'UNI che per la parte ICT si avvale dell'UNINFO.

SUMMARY

The evolution of ICT not only pays attention to the traditional process quality, but to the quality of products too, in order to stress the users' needs. For this reason, in order to verify the strictness of the information, have been introduced control tests and quality standard models which should be followed. These models are based mainly on new aspects related to requirements, assessment, management and quality measurement. For this reason has been developed the standards SQuaRE, which reflects the great international activities and in which INAIL has contributed. In Italy the standardization activity is followed by UNI supported by UNINFO for ICT.

BIBLIOGRAFIA

- [1] DONALD A., NORMAN: *The psychology of every day things*, 1988.
- [2] D. NATALE: *On the impact of metrics application in a large software maintenance environment*, IEEE, Conference on Software Maintenance, CSM'91, Sorrento, 1991.
- [3] G. MANTOVANI: *L'interazione Uomo-computer*, Il Mulino, 1995.
- [4] D. NATALE: *Qualità e quantità nei sistemi software*, FrancoAngeli, 1995.
- [5] *Quality in the European Statistical System - The Way Forward*, 2002 Edition, (Leg on Quality), Luxembourg.
http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ESS_QUALITY_RECOMMENDATIONS_2002_EN_0_1.pdf.
- [6] *Definition of quality in statistics*, Working group *Assessment of quality in statistics*, Luxembourg, 2-3 October 2003.
<http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/ess%20quality%20definition.pdf>.
- [7] *Standard Quality Report*, Methodological Documents, Working Group "Assessment of quality in statistics", Luxembourg, 2-3 October 2003.

http://epp.eurostat.ec.europa.eu/portal/page/portal/quality/documents/STANDARD_QUALITY_REPORT_0.pdf

[8] *ISO/IEC 25000, Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE*, 2005.

[9] C. BATINI, M. SCANNAPIECO, *Data Quality*, Springer, 2006.

[10] ISO/IEC 25012, SOFTWARE ENGINEERING - SOFTWARE PRODUCT QUALITY, Requirements and Evaluation (SQuaRE) - Data Quality Model, 2008.

[11] D. NATALE: *La qualità dei dati e l'ISO/IEC 25012*, Rivista *Unificazione&Certificazione*, UNI, Milano, 2009.

[12] Tratte da Eurostat (2009) *ESS Handbook for Quality Reports*, 2009 Edition, EUROSTAT: Methodologies and working papers, http://epp.eurostat.ec.europa.eu/portal/page/portal/lang-en/ver-1/quality/documents/ESQR_FINAL.pdf

[13] D. NATALE: *ISO/IEC Modelo de Calidad de datos y Data Governance*, Capitulo 3, in C. Calero, MA. A. MORAGA, M. G. PATINI, *Calidad del Producto y proceso software*, Ra-Ma, Madrid, 2010.

[14] ISO/IEC 25010, SOFTWARE ENGINEERING - SOFTWARE PRODUCT QUALITY REQUIREMENTS AND EVALUATION (SQUARE) - Software and system quality model, 2011.

[15] D. NATALE: *Complexity and data quality*, Poster e Atti Conferenza 13-16 settembre, CHItaly 2011, Alghero, 2011.

[16] D. NATALE: *Data quality e Open Data*, *Convegno AICA 2011*, Torino, 2011.

[17] PASQUALOTTO E., FEDERICI S., SIMONETTA A. & OLIVETTI BELARDINELLI M.: *Usability of Brain Computer Interfaces*.

Association for the Advancement of Assistive Technology in Europe (AAATE 2011), Assistive Technology research series, Maastricht, Netherlands, Aug 31- Sep 02, 2011. In G. J. GELDERBLOM, M. SOEDE, L. ADRIAENS, & K. MIESENBERGER (Eds.), *Assistive Technology Research Series: Vol. 29. Everyday technology for independence and care*. (pp. 481-8). Amsterdam: IOS Press. doi:10.3233/978-1-60750-814-4-481.

[18] CISCO VISUAL NETWORKING INDEX: *Forecast and Methodology*, 2010-2015, CISCO white paper. Giugno, 2011.

[19] *IDC Digital Universal Studym*, sponsored by EMC, May 2010.

[20] SIMONETTA A.: *Nuovi scenari e prospettive della Brain Computer Interface (BCI)*. *Rivista sugli infortuni e le malattie professionali* - IV Serie della rassegna della previdenza sociale anno XCVIII. Fascicolo 2/2011.

[21] NWI ISO/IEC 25022, SOFTWARE ENGINEERING, Software Product Quality Requirements and Evaluation (SQuaRE), Measurement of quality in use, 2012.

[22] NWI ISO/IEC 25023, SOFTWARE ENGINEERING, Software Product Quality Requirements and Evaluation (SQuaRE) - Measurement of system and software product quality, 2012.

[23] NWI ISO/IEC 25024, SOFTWARE ENGINEERING, Software Product Quality Requirements and Evaluation (SQuaRE), Measurement of data quality, 2012.