

Applicazione di procedure ISO per la valutazione e documentazione della qualità dei geodati

Daniela Carrion (*), Andrea Maffeis (*), Federica Migliaccio (*)

(*) DIIAR sez. Rilevamento – Politecnico di Milano, Piazza Leonardo da Vinci 32 – 20133 Milano
Tel. 02 23996502 – fax 02 23996530

Riassunto

Lo scopo di questo lavoro è stato quello di indagare le ragioni fondanti del tema della qualità dei geodati con riferimento agli standard adottati in Europa ed alla loro applicabilità, volendo anche sperimentare delle procedure per la valutazione della qualità dei dati.

Il lavoro si è svolto su più fronti: dapprima si sono analizzati nel dettaglio i tre standard ISO (19113, 19114, 19115) che propongono un quadro di riferimento per la descrizione – valutazione – documentazione della qualità dei dati e che sono stati recentemente (2005) recepiti anche come Standard Italiani; di conseguenza, al fine di conoscere l'effettiva diffusione ed applicazione di questi tre standard ISO a livello italiano, è stata effettuata un'indagine (anche sulla base di analoghe esperienze a livello europeo) tra gli uffici cartografici delle regioni e delle province italiane. Infine, è stata eseguita una valutazione della qualità per dei dataset reali documentando i risultati ottenuti seguendo le indicazioni di ISO 19113, ISO 19114 e ISO 19115.

Abstract

The aim of this paper was to research the basic reasons laying the basis of spatial data quality, with reference to the European Standards and their usability.

The task was carried out by several lines: at first the three ISO Standards (19113, 19114, 19115) were analyzed in detail. The Standards propose a complete framework for the description, the evaluation and the reporting of spatial data quality. They have been recently (2005) acknowledged as official Italian Standards too.

Then, a questionnaire addressed to cartographic offices of the Italian regions and provinces was set up and distributed in order to know the real use and spreading of the three Standards in the Italian contest.

Finally a quality evaluation according to ISO 19113, ISO 19114 and ISO 19115 was experimented on real datasets and the obtained results were reported as metadata.

Indagine conoscitiva sull'uso degli standard ISO per la qualità nella geoinformazione

Il panorama dell'Informazione Geografica ha subito negli ultimi anni una notevole spinta in termini di disponibilità di dati, applicazioni e sistemi. A fronte di ciò, la questione della qualità dei dati e delle informazioni che confluiscono nei GIS è rimasta generalmente a margine, assumendo solo ora un ruolo chiave. Su questi temi si veda ad esempio Carrion e Migliaccio (2005), dove viene indicato il quadro di riferimento delineato dagli Standard ISO 19113, 19114, e 19115 che tracciano le linee guida per la definizione e valutazione della qualità dei dati in un Sistema Informativo Geografico.

Gli standard ISO relativi alla valutazione di qualità ed alla documentazione (ISO 19113, 19114, 19115) sono stati solo recentemente recepiti a livello italiano come standard ufficiali quindi non esisteva alcun tipo d'informazione circa la loro reale diffusione ed utilizzo negli uffici cartografici delle regioni e delle province italiane.

Num. ISO	Nome	Anno	UNI
ISO 19113	Geographic Information – Quality principles	2002	UNI EN ISO 19113:2005
ISO 19114	Geographic Information – Quality evaluation procedures	2003	UNI EN ISO 19114:2005
ISO 19115	Geographic Information – Metadata	2003	UNI EN ISO 19115:2005

Figura 1: Standard relativi alla qualità nella geoinformazione.

Partendo da questo presupposto è maturata l'idea di effettuare un'indagine conoscitiva, anche sulla base di analoghe esperienze a livello europeo. L'indagine è stata condotta attraverso l'invio, nel dicembre 2005, di un apposito questionario agli uffici cartografici delle regioni e delle province italiane. L'indagine si è conclusa nell'aprile del 2006 avendo ricevuto 24 questionari compilati, corrispondenti a più di un terzo degli enti contattati.

I risultati, riportati nelle figure sottostanti, mostrano che la diffusione degli standard ISO non è ancora sistematica e capillare, emerge inoltre la mancanza di comportamenti univoci da parte degli enti contattati, che pure dimostrano grande interessamento e buona volontà. Vediamo in breve i risultati emersi. La totalità degli Enti intervistati è dotata di un SIT, e di questi circa l'80% è pubblicamente consultabile da Internet. Una domanda del questionario era incentrata sulla disponibilità di metadati per queste geoinformazioni (da notare che non veniva richiesta l'adesione ad un particolare standard). I risultati hanno indicato che meno del 10% degli enti contattati possiede una documentazione completa e disponibile dal Web. D'altro canto gli enti che dichiarano di non possedere nessun tipo di metadato sono poco più del 10% (si veda la Figura 2).

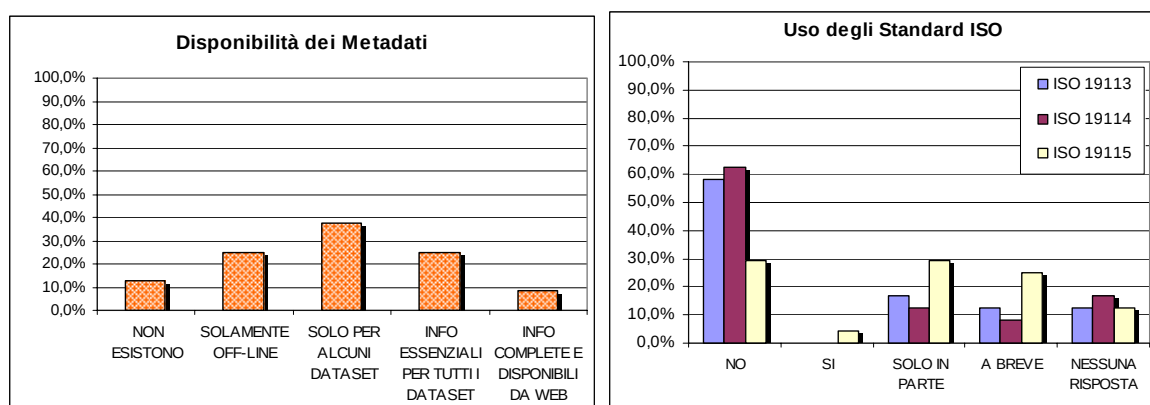


Figura 2 - Disponibilità di Metadati per i WebSIT degli Enti contattati e utilizzo degli Standard ISO per la valutazione e descrizione della qualità dei dati

Per quanto riguarda l'utilizzo degli standard ISO 19113, 19114 e 19115 relativi alla valutazione e alla documentazione della qualità dei geodeti, è interessante notare come la loro diffusione presso gli uffici cartografici italiani non vada di pari passo, bensì si noti una maggiore predisposizione all'utilizzo della norma ISO 19115 rispetto alle altre due. Infatti per ISO 19113 e ISO 19114 gli uffici cartografici che dichiarano di non utilizzarli per nulla sono tra il 55% ed il 60%, mentre per ISO 19115 tale valore scende al 29,2% (si veda sempre la Figura 2).

La totalità degli enti intervistati considera la tematica della qualità per la geoinformazione all'interno di un SIT di importanza strategica, ma dal punto di vista pratico, solamente il 28% degli enti è dotato di un sistema per la descrizione – valutazione – documentazione della qualità dei dati. Questo indica che esiste una significativa contrapposizione tra ciò che viene ritenuto importante e ciò che poi viene effettivamente fatto.

Infine è significativo riportare le opinioni sull'argomento degli addetti ai lavori: dall'indagine è emerso chiaramente che gli Standard sono largamente visti come un'opportunità interessante, ma

ancora di difficile applicazione. Quando è stato chiesto invece di indicare quali siano le ragioni di tale problematicità, quello che risulta è la mancanza, da una parte di competenze e risorse, e dall'altra di adeguati comportamenti istituzionali.

Sperimentazione sull'applicazione degli standard ISO

La sperimentazione ha riguardato l'applicazione delle procedure dagli standard ISO per la descrizione, valutazione e documentazione della qualità dei geodati, con l'obiettivo di verificare il comportamento delle specifiche ISO, nel momento in cui esse devono essere implementate per la valutazione di dataset reali.

Anche se nello standard ISO 19114 sono indicati i passi della valutazione e sono forniti alcuni esempi applicativi, nella pratica per ogni situazione particolare ci si scontra con problematiche non banali, che se non analizzate debitamente possono influenzare la bontà del processo di valutazione.

I dataset scelti per la sperimentazione sono stati estratti dalla base di dati che costituiva il prototipo del SIT del Comune di Montichiari (BS), sviluppato dalla sezione Rilevamento del DIIAR – Politecnico di Milano negli anni scorsi. Il prototipo era stato realizzato solo per una parte del territorio comunale e comprendeva non solo i dataset frutto della trasposizione da cartografia numerica a database topografico ma anche un principio di integrazione tra le mappe catastali e la Carta Tecnica Comunale, nonché altri dati di tipo non georeferenziato. Gli elementi di qualità valutati durante la sperimentazione sono: accuratezza posizionale assoluta, completezza e consistenza logica e saranno analizzati nel dettaglio nei prossimi paragrafi.

Per ogni "oggetto di qualità" si sono trovate delle misure significative (con specifiche unità di misura correlate) in grado di valutare la qualità dei dataset in riferimento all'elemento voluto. Proprio per questo punto, in cui la scelta di una misura piuttosto che un'altra può risultare critica, la norma ISO 19114 si limita a indicare il formalismo della struttura e a fornire qualche esempio. Sarebbe di certo più utile fornire dei criteri o delle indicazioni su cui basarsi per la scelta.

Per quanto riguarda l'accuratezza posizionale, la misura di qualità scelta è stata la media campionaria dello scostamento tra punti omologhi, mentre per la completezza sono stati conteggiati il numero di elementi immessi e quelli omessi.

Per completare il quadro della descrizione – valutazione – documentazione, ottenuti i risultati di qualità dei geodati, sono stati compilati i metadati (solamente la parte relativa al pacchetto "qualità dei dati") per ogni dataset valutato e per alcuni di questi si è redatto anche un report di qualità.

Valutazione dell'accuratezza posizionale assoluta secondo gli standard ISO

Per quanto riguarda i Sistemi Informativi Territoriali il concetto di accuratezza posizionale ricopre un'accezione particolare. Infatti in questi sistemi è usuale integrare dataset di natura diversa e provenienti da fonti diverse, per i quali la sovrapposizione non è assicurata.

Nel caso in esame del Comune di Montichiari, il modello concettuale indicava, tra gli strati informativi da integrare nel SIT, la carta catastale e la carta tecnica comunale (su cui sono appoggiati anche gli altri strati):

- la carta catastale deriva da una recente opera di informatizzazione del patrimonio mappale di competenza del Comune di Montichiari: in pratica sono stati presi i fogli catastali originali, sono stati scannerizzati e digitalizzati;
- la carta tecnica comunale alla scala 1:2000 è stata prodotta recentemente (2003), derivandola da un rilievo aerofotogrammetrico del 2001.

Quindi è assolutamente naturale che il contenuto posizionale degli elementi di questi due strati informativi non sia perfettamente identico.

Stando così le cose, è naturalmente importante avere un'idea di quanto un dataset (ad esempio la mappa catastale) si discosti dall'altro (aerofotogrammetrico) sia che questa distanza rappresenti il punto di partenza sia che rappresenti il punto d'arrivo. Tale ragionamento ha ispirato la valutazione dell'accuratezza posizionale effettuata in questo studio di caso.

Nella fase di progettazione del S.I.T. di Montichiari svolta negli anni passati, al fine di integrare i due strati informativi (Catastale e Aerofotogrammetrico), era stato scelto un certo numero di coppie di punti omologhi (attraverso una procedura automatica) e poi erano state eseguite delle trasformazioni con metodi differenti e con una diversa densità di coppie di punti (11250 e 4276). Il processo di valutazione è stato applicato ai due dataset disponibili. Per ognuno dei due dataset si sono calcolate le distanze fra punti omologhi. Tali distanze forniscono una misura di accuratezza posizionale del dataset fornito dai punti della carta catastale, usando la carta tecnica comunale come dato di confronto. I risultati delle prove di accuratezza posizionale sono riassunti nella Figura 3 per entrambi i dataset valutati.

<i>Tema</i>	<i>Vettori_4276.shp</i>		<i>Vettori_11250.shp</i>
N° elementi	4276		11248
MEDIA	0,846 m		1,533 m
DEV.ST.	0,399 m		0,768 m
MAX	2,107 m		3,784 m
MIN	0,025 m		0,034 m

Figura 3: Risultati dell'accuratezza posizionale per i due dataset di 4276 e 11250 punti

I risultati ottenuti mostrano chiaramente differenze fra lo scarto quadratico medio tra le coppie di punti omologhi del dataset con 4276 punti e quello con 11250 punti: a parte le possibili considerazioni da fare sulle cause che possono aver generato questo risultato, nell'ambito della valutazione di qualità tale problema ha ripercussioni sulla determinazione dell'affidabilità della misura di accuratezza, punto non banale da risolvere e che dovrà essere oggetto di future ulteriori riflessioni.

Valutazione della completezza secondo gli standard ISO

In generale, l'obiettivo delle valutazioni di completezza per un dataset geografico (o per un suo sottoinsieme o una serie di dataset) è la verifica della presenza nel dataset di tutti gli elementi descritti nell'Universo del Discorso¹ (UdD) e della mancanza di elementi aggiuntivi. Per la valutazione della completezza il termine di confronto è rappresentato dal modello concettuale, oppure da un altro dataset, per cui sia però noto il valore di qualità.

Le norme ISO scorrono l'elemento di completezza in due sottoelementi: omissione di un elemento e immissione nel dataset di un elemento estraneo. La misura di completezza di un dataset può essere espressa in termini assoluti (numero di elementi mancanti/immessi) in termini percentuali o in termini di conformità (nel caso sia fornito un valore da rispettare). Inoltre, sebbene sia possibile compiere la verifica di completezza, in linea teorica, per ogni elemento del dataset, nella pratica, essendo molto difficoltoso operare dei controlli automatici, spesso si deve ricorrere a dei controlli a campione.

Per quanto riguarda il SIT di Montichiari, era stato redatto un capitolato per la trasposizione della cartografia numerica verso dei formati compatibili con un GIS. Il lavoro di "restituzione" era stato affidato, in via preliminare, a due aziende del settore, che avevano prodotto per un'area campione gli shapefile di vari tematismi, organizzati secondo la struttura e la tipologia indicata dal capitolato. Vista la possibilità di disporre di due dataset indipendenti che si riferivano alla stessa modellizzazione del territorio e visto che tali restituzioni presentavano delle sensibili differenze, si è effettuata la simulazione della valutazione di completezza per questi due dataset.

E' evidente che l'obiettivo non era valutare quale delle due ditte (che nel seguito chiameremo A e B) avesse lavorato meglio, ma capire come si possa operare una valutazione della qualità dei dati.

¹ Universo del Discorso: Termine comunemente usato in letteratura per riferirsi al processo di modellizzazione concettuale delle entità nel mondo reale.

Invece che campionare i due dataset per cercare i rispettivi errori, si è preferito evidenziare gli elementi che differivano nelle due restituzioni, generando quindi per ogni strato informativo uno shapefile contenente unicamente gli elementi “differenti”, e verificarne manualmente la correttezza in riferimento ai dati di partenza effettuando un confronto incrociato tra l’elemento stesso, il file AutoCad originale ed il capitolato, per stabilire anche la causa della differente interpretazione. L’identificazione delle differenze tra i due dataset è stata fatta per tutti gli strati informativi.

L’elemento che differisce nei due dataset può essere infatti il frutto di un errore di omissione da parte di A oppure da parte di B, o di un errore di immissione da parte di A oppure da parte di B; ancora, la differenza potrebbe essere solamente il risultato di una diversa applicazione delle specifiche, senza peraltro essere in contraddizione con queste.

Di seguito si riportano i risultati finali per i tematismi “tronco stradale” ed “aiuole urbane”, per i quali sono emerse le maggiori differenze (si veda la Figura 4).

<i>Strato informativo e Ditta</i>	<i>N° totale elementi</i>	<i>N° elementi omessi</i>	<i>N° elementi immessi</i>
<i>Tronco stradale (A)</i>	27	9	3
<i>Tronco stradale (B)</i>	68	1	1
<i>Aiuole (A)</i>	8	0	0
<i>Aiuole (B)</i>	11	0	3

Figura 4: Risultati della valutazione di completezza

Valutazione della consistenza logica secondo gli standard ISO

I dataset utilizzati sono sempre quelli prodotti dalle due aziende A e B per l’area del prototipo citato precedentemente. Secondo quanto indicato dal modello concettuale esplicitato nel capitolato, a partire dalla cartografia numerica in formato AutoCAD e dai codici di classificazione degli elementi, si sarebbe dovuto produrre un dataset costituito da una serie di tematismi (ognuno dei quali rappresentato con le necessarie tipologie geometriche) all’interno dei quali ci fossero i layer corrispondenti agli strati informativi. In aggiunta al capitolato, era stata prodotta anche una tabella che per ogni strato informativo, indicava lo stato di fatto della CN e la corretta geometria dei singoli layer coerentemente con quanto indicato dal capitolato.

Ad una prima analisi risultava che le specifiche del capitolato non sempre erano state rispettate, quindi la valutazione ha riguardato la consistenza di formato tra i layer restituiti e le specifiche stesse. Operativamente, è stata costruita una tabella in cui erano elencati tutti i layer e gli strati informativi contenuti nel capitolato, e per ogni layer si è verificata l’effettiva geometria nei dataset restituiti dalle due ditte. Per ogni strato informativo, si è calcolato il numero di layer consistenti nella trasposizione di ogni ditta, così come emerge dalla Figura 5.

La misura di qualità scelta per la serie di dataset è il rapporto tra layer consistenti e il numero totale di layer.

<i>Strato informativo</i>	<i>n° layer</i>	<i>layer coerenti</i>	
		<i>Ditta A</i>	<i>Ditta B</i>
Infrastrutture e manufatti stradali	12	5	12
Costruzioni	45	36	43
Idrografia, impianti e manufatti	12	9	12
Vegetazione	8	7	7
Orografia e punti di coordinate note	11	9	11
<i>Totale</i>	<i>88</i>	<i>75 %</i>	<i>97%</i>

Figura 5: Risultati della valutazione di consistenza logica

Conclusioni

Nel complesso, ad oggi, dal punto di vista di una capillare diffusione ed utilizzo degli standard ISO per la descrizione – valutazione – documentazione della qualità dei dati, la situazione italiana si presenta piuttosto carente e lacunosa. A fronte di una grande quantità di dataset ed informazioni disponibili nei vari campi d'interesse dei Sistemi Informativi Territoriali da parte delle amministrazioni locali, la documentazione della qualità dei geodati ricopre un ruolo marginale. Nonostante ciò, un buon numero di addetti ai lavori dimostra forte interessamento e una particolare sensibilità verso la tematica della normazione nella GI in generale e della qualità dei dati in particolare, segnale evidente di una situazione in fermento. Tale interessamento dimostrato dai singoli è molto importante, ma non può essere sufficiente.

Sarebbe molto importante che in questo campo gli standard si svincolassero da quella visione puramente formale che sembrano avere oggi. Se l'applicazione degli standard risulta essere troppo macchinosa, ingiustificatamente formalista e senza reali ricadute nella gestione dei dataset territoriali, difficilmente si potranno vedere i frutti del loro utilizzo.

Per quanto riguarda la sperimentazione è stato possibile evidenziare alcuni punti critici del processo di valutazione di qualità, in particolare:

- la scelta di un'adeguata misura di qualità;
- la disponibilità di un riferimento certo con cui operare il confronto;
- l'attuale mancanza di validi strumenti per la compilazione dei metadati coerenti con lo Standard ISO 19115 e le recenti specifiche nazionali.

In conclusione, dato per assodato che i riferimenti normativi per l'Informazione Geografica sono gli standard ISO, sarebbe altresì necessario in ambito nazionale fornire a tutti i soggetti interessati un supporto all'implementazione degli Standard (come delle linee guida o altro) che siano un concreto aiuto e riferimento, non solo per quanto riguarda il rispetto delle specifiche formali, ma anche dal punto di vista dei contenuti.

Bibliografia

Burrough P., McDonnell A. (1998), *Principles of Geographical Information Systems*, Oxford University Press, Oxford

Caprioli M., Tarantino E. (2003), "Standards and Quality in GIS contexts", *Proceedings of FIG Working Week 2003*

Carrion D., Migliaccio F. (2005), "Importanza degli Standard nella valutazione della qualità dei dati in un Sistema Informativo Territoriale", *Atti della 9ª conferenza nazionale ASITA*, Catania, 15 – 18 Novembre 2005, Vol. I, 619 - 624

Jakobsson A., Wasström C. (2004), "Use of the ISO 19100 Quality standards at the NMCAs – Results from a questionnaire", www.eurogeographics.org

Mood A.M., Graybill F.A., Boes D.C. (1974), *Introduction to the theory of statistics*, McGraw – Hill, New York

Shi W., Fisher P.F., Goodchild M.F. (2002), *Spatial data quality*, Taylor & Francis, London & New York

UNI EN ISO 19113:2005 – Quality Principles

UNI EN ISO 19114:2005 – Quality evaluation

UNI EN ISO 19115:2005 – Metadata

Ringraziamenti

La presente nota è stata redatta nell'ambito dei lavori relativi alla ricerca COFIN 2004 dal titolo "Strutture evolute della cartografia numerica per i GIS e l'ambiente WEB", coordinatore scientifico Prof. R. Galetto, Unità di Ricerca del Politecnico di Milano.

Si ringrazia inoltre il Comune di Montichiari per aver reso disponibili i dati per questa ricerca.

Si ringrazia infine il Laboratorio di Geomatica del Polo di Como del Politecnico di Milano per l'estrazione del campione di punti omologhi.