Il DataBase di Sintesi Nazionale (DBSN) - Banca dati geografica e base cartografica digitale

Carlo Perugi[(1)] Cinzia Tafi[(2)]

[(1)] Istituto Geografico Militare, Direzione Acquisizione Dati, via Cesare Battisti 10, 50122 Firenze
vdiracdati@geomil.esercito.difesa.it
[(2)] Istituto Geografico Militare, Direzione Acquisizione Dati, via Cesare Battisti 10, 50122 Firenze
caservgdbn@geomil.esercito.difesa.it

Abstract
L'IGM produce il DBSN (DataBase di Sintesi Nazionale) a copertura nazionale. I dati permettono di analizzare il territorio e di creare rappresentazioni cartografiche a varie scale. L'utilizzo del DBSN ha il vantaggio di gestire una base di dati in continuo geografico a livello nazionale mantenendo il livello di dettaglio regionale. Per raggiungere tale obiettivo si è ricorso alla raccolta di dati territoriali prodotti da Enti pubblici e da Open Street Map (OSM) da validare e uniformare a livello nazionale. Attualmente la banca dati è in fase di ulteriore revisione con interventi correttivi e di integrazione per risolvere criticità e incompletezze presenti nei dati originari.

Parole chiave: Database, Geotopografico, Qualità, Cartografia.

1 Il progetto: utilità del database di valenza nazionale

Il DBSN, in quanto database omogeneo su tutto il territorio nazionale, costituito da 9 strati tematici geotopografici che interessano la base informativa di interesse comune con livello di dettaglio regionale, permette di analizzare il territorio relazionando dati di copertura del suolo con viabilità, o con l'edificato e l'idrografia, al fine di pianificare interventi sul territorio sia per progettazione di nuove infrastrutture sia per valutazioni di rischi territoriali anche in conseguenza di situazioni di emergenza (soprattutto quando occorre l'informazione su aree che investono più regioni).
Di seguito si descrivono alcuni esempi di utilizzo dei dati.

In fig. 1, in rosso sono evidenziati gli edifici sedi di scuole statali e paritarie dall'infanzia all'istruzione secondaria di secondo grado.

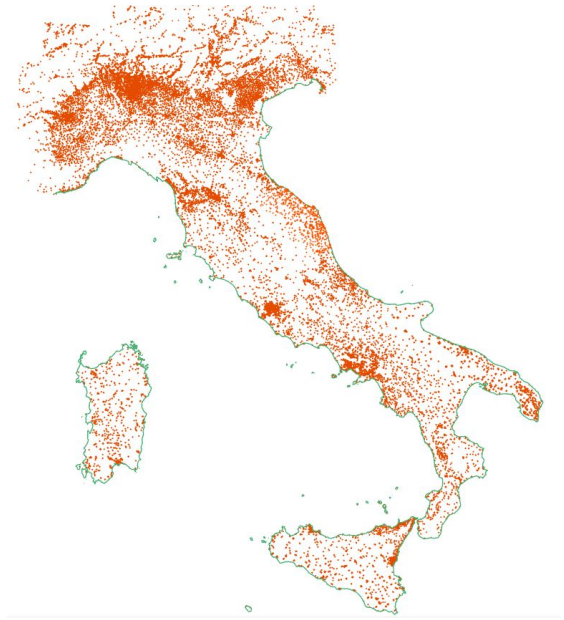

Fig.1. Edifici scolastici sul territorio nazionale.

Al di là delle semplici informazioni relative al numero totale dei plessi scolastici ed ai metri quadrati totali occupati dagli edifici ospitanti, è interessante derivare ulteriori informazioni mettendo in relazione dati appartenenti a classi diverse.
Può essere utile, per esempio, confrontare su base comunale il numero degli istituti scolastici con la popolazione residente e creare così una statistica comparativa che evidenza a livello nazionale dove esiste una maggiore esiguità di strutture rispetto alla domanda legata al numero dei residenti.

Altro esempio potrebbe essere dato da i parametri di pericolosità per rischio idrogeologico che sono calcolati come chilometri quadrati per ogni comune, confrontati con la presenza di aree industriali o aree legate all'industria culturale quali siti archeologici, musei o manufatti religiosi di alto valore storico, pievi e conventi.
Anche in questo caso si possono trarre indicazioni utili per le scelte da operare, a livello nazionale, nell'indirizzare eventuali risorse da impiegare.

La possibilità di svolgere indagini comparative di ogni tipo su tutto il territorio nazionale costituisce un valore aggiunto rispetto ad una massa di dati enorme, non sempre condivisibili e corrette, spesso presi in considerazione solamente nel loro ambito tematico.
E' evidente come sia indispensabile avere a disposizione, per le molteplici interrogazioni possibili, dati aggiornati e con un livello di completezza che tende asintoticamente al 100% e questo pone il problema di come sia indispensabile dotarsi di strumenti validi di verifica e validazione dei dati per l'attendibilità dei risultati.

2 Primo impianto del DBSN

Fig. 2. Composizione dei dataset regionali per la formazione del DBSN.

La gran massa dei dati che costituisce il DBSN proviene dai dataset regionali (Fig. 2) confluiti in un unico schema, resi quindi omogenei a livello di struttura ma non per densità ed aggiornamento di informazione. Per quest'ultimo scopo è necessario un lavoro di integrazione condotto con dati ufficiali e certificati come ad esempio Istat e Catasto e dati non ufficiali come Open Street Map ma utili per il livello di aggiornamento e la ricchezza di informazione. Integrare significa acquisire occorrenze complete del contributo valoriale degli attributi o trasferire solamente il valore degli attributi a geometrie esistenti ritenute valide. Un primo ciclo di attività in tal senso ha coperto un arco temporale di tre anni riguardando una selezione di classi prioritarie mentre attualmente siamo circa a metà di un secondo ciclo in cui si cerca oltre alle operazioni di integrazione, di affrontare altri aspetti del Database descritti nei paragrafi successivi.

3 Operazioni per ottimizzare il DBSN

3.1 Classi di importanza nazionale

Al fine di migliorare la gestione del database e organizzare il lavoro di integrazione sono state applicate alcune operazioni di seguito descritte per le lavorazioni. In particolare sono state applicate procedure e integrazioni su due livelli di database che compongono il DBSN: database regionale e database nazionale, quest'ultimo contenente le classi che preferibilmente sono da lavorare per tutto il territorio nazionale non essendo molto numerose ed avendo fonti di dati originarie già nazionali. Tali classi definite "di importanza nazionale"non vengono editate regione per regione ma sono riconducibili alle singole regioni in base all'attributo *cod_reg* (codice regionale) popolato per ogni istanza.
La classi interessate sono le seguenti:

- Acqua interna (*acq_in*)
- Acqua territoriale (*acq_ter*)
- Aree naturali protette (*ar_nat*)
- Vertice di rete (*v_rete*)
- Tin, modello digitale del terreno, (*z_tin*)
- Caposaldo (*caposd*)
- Comune (*comune*)
- Linea di costa marina cartografica (*cs_mar*)
- Curva di livello (*cv_liv*)
- Elemento di trasporto su acqua (*el_acq*)
- Ghiacciaio-nevaio perenne (*ghi_nv*)
- Limiti amministrativi (*lm_am*)
- Curva batimetrica (*ln_btm*)
- Località significative (*loc_sg*)

Provincia (*provin*)
Punto batimetrico (*pt_btm*)
Punto quotato (*pt_quo*)
Regione (*region*)
Stato (*stato*)

Nelle classi *region, provin, comune* e *ln_am* sono stati inseriti le features derivate dall'elaborazione dei limiti amministrativi comunali mediante procedure sviluppate internamente. Precisamente i limiti comunali sono stati prodotti in automatico dopo risoluzione problemi di errate georeferenzazioni e rendendo le aree e le linee topologicamente congruenti fra sé, con il confine di stato e con la linea di costa. I dati iniziali sono sono quelli ISTAT del 2021 ma i confini prodotti a valle dell'elaborazione non costituiscono un dato ufficiale salvo il confine di stato dell'IGM (prodotto nel 2016 e con alcuni aggiornamenti del 2022).

Per la la classe *el_acq* sono stati inseriti i dati di Eurograographics; per questa classe non è assegnato l'attributo *cod_reg* tranne che per le vie lacuali e fluviali.
La classe lineare *lm_am* non rientra nella specifica di contenuto del DBSN ma è stata creata per scopi cartografici e contiene l'attributo dei limiti con indicato il comune a destra e sinistra della linea.
Le classi *acq_ter* e *acq_in* (acque territoriali ed interne) sono state perimetrate a partire dalla loro definizione nelle gazzette ufficiali di riferimento.

3.2 Valori di default per gli attributi

Per quanto riguarda il riempimento degli attributi di tutte le classi del DBSN si procede per ogni regione che viene lavorata all'eliminazione di valori nulli inserendo valori di default in base ad una tabella predefinita. L'inserimento in alcuni casi è soggetto all'applicazione di query o tenendo conto di associazioni tra classi per rendere congruenti le attribuzioni delle classi. Ad esempio per l'idrografia la tipologia dell'elemento idrico (*el_idr_ty*) occorre applicare la query per trovare i tratti che attraversano le dighe ed i laghi/invasi, per questi tratti è previsto il tipo 02-virtuale.

3.3 Elaborazione dei nodi delle reti

Le classi puntuali corrispondenti ai nodi delle reti sono considerati elementi critici da elaborare a parte in quanto se popolati con i dati originari regionali non sempre coincidono con i tratti lineari delle reti di appartenenza. Molti sono geometrie logiche di cambio attributo o nodi terminali, in alcuni casi rappresentano oggetti realii da verificare come ad esempio i passaggi a livello della rete ferroviaria, oggetti comunque scelti perché di interesse.
In molti casi gli interventi portano ad eliminare i tratti ed i nodi restano isolati. In altri casi gli oggetti rappresentati dai nodi non sono più esistenti.
Per questi motivi si opera una cancellazione dei nodi di carattere logico per ricrearli successivamente con procedura automatica in coincidenza topologica con gli elementi lineari.

Per la rete idrica vengono eliminati i nodi idrici (*nd_idr*) che collegano gli elementi idrici (*el_idr*), soggetti alla fase di integrazione/correzione; a fine lavorazione, per ogni regione, si ricreano i nodi.
Per alcuni nodi che rappresentano oggetti si effettua un controllo dell'elemento selezionando solo quelli di interesse.

Per la rete stradale dell'Intersezione stradale (*iz_str*) si lasciano solo quelli con tipologia ty = 02 (casello autostradale) eliminando tutti gli altri.

Per la rete ferroviaria consideriamo la Giunzione ferroviaria (*gz_fer*) solo per il tipo ty=01 (passaggio a livello) e ty=04 (stazione/fermata/casello), eliminando tutti gli altri.

Per la rete degli elettrodotti si considera la classe Sostegno a traliccio (*tralic*) come oggetto correlato alle linee elettriche togliendo i nodi della rete elettrica (*nd_ele*) ma lasciando solo quelli in corrispondenza delle sottostazioni elettriche o cabine di trasformazione.

3.4 Analisi dati e rilevazioni errori per fase di integrazione

Nella prima fase di integrazione/correzione l'attività si è concentrata sugli oggetti ritenuti prioritari. Ci si è basati prevalentemente su una lista di "priorità 1" comprendente la viabilità principale, l'idrografia, l'edificato riguardante edifici significativi (municipi, chiese, scuole, industrie, musei).

Nella seconda fase di intervento, avviata a partire dalle regioni del sud Italia, sono stati presi in considerazione ulteriori oggetti, lista di "priorità 2" comprendente alcune tipologie di viabilità secondaria, di manufatti industriali, di pertinenze e di edifici minori.

Ai fini della pianificazione di questa seconda fase di integrazione/correzione dei dati si effettua un'analisi di dettaglio per individuare anomalie geometriche ed incongruenze di classifica. In genere sono caratteristiche di ogni regione in base al tipo di acquisizione dei dati originari.

Alcuni errori sono risolvibili in nodo sistematico su tutta la regione attraverso la costruzione di procedure in model builder (in ambiente ESRI) o sviluppate in Python. Di seguito alcuni esempi.

Fig. 3. Accorpamento edifici.

ACCORPAMENTI EDIFICI

Errore ricorrente (Fig. 3) è quello degli edifici non suddivisi per categoria d'uso o per tipologia: ci sono corpi edificati di grandi dimensioni comprendenti alcuni edifici di interesse prioritario come ad esempio le chiese. In tal caso si interviene nel database suddividendo l'edificio e cambiando la classifica della porzione appartenente alla chiesa.

GEOMETRIE SOVRAPPOSTE

In altri casi (Fig. 4) ci sono errori creati in fase di trasformazione di schema dai dati originari alla struttura del DBSN. In particolare dove elementi puntuali nel DBSN devono assumere geometrie areali sono state create aree fittizie con utilizzo di funzioni buffer in modo automatico, come per esempio silos o ciminiere tra i manufatti industriali. Questo comporta che gli stessi elementi risulteranno sovrapposti tra loro e con altri oggetti limitrofi per cui occorrono interventi per operare spostamenti e/o eliminazione di alcuni elementi.

Fig. 4. Sovrapposizione di geometrie.

ERRORI DI CLASSIFICA CLASSE

In altri casi (Fig. 5) ci sono classifiche di oggetti totalmente errate, come per esempio le condotte classificate come edifici, caso individuabile sulla base delle dimensione e delle forme.
Altro esempio è la classifica di edifici ed edifici minori in base alla loro struttura: alcuni edifici corrispondono a tettoie o baracche estese, ben visibili sulle ortofoto Agea e sono invece classificate edifici.

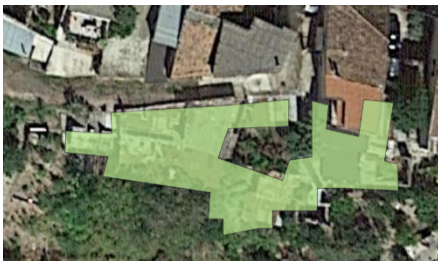

Fig. 5. Classificazione errata.

Fig. 6. Aggiornamento da considerare.

AGGIORNAMENTI

Il controllo del contenuto del database (Fig. 6) basato sul confronto di ortofoto Agea aggiornate, permette di evidenziare la necessità di aggiornare la classifica di alcuni oggetti o l'eliminazione degli stessi. Ad esempio un campo sportivo abbandonato.

DISOMOGENEITA' DI DENSITA'

In alcune regioni sono evidenziate delle disomogeneità nella densità degli oggetti (Fig. 7), in particolare nell'idrografia, dovute alle modalità di acquisizione dei dati originari.

In queste situazioni si interviene cercando di colmare le informazioni nelle aree meno dense di elementi.

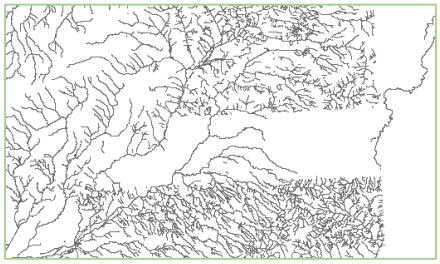

Fig. 7. Densità disomogenea.

4 Valutazione di qualità e reportistica

Successivamente a questa ulteriore integrazione e correzione dei dati si procede alla valutazione di qualità del DBSN procedendo per regioni. La valutazione di qualità segue i criteri standard ISO/TC211 (International Standard Organization – Geographic Information/Geomatics) e, in particolare, nel documento ISO 19157 (Geographic Information – Data Quality) a cui si far riferimento per la determinazione dei parametri di conformità di struttura alle specifiche di contenuto, completezza, accuratezza posizionale e tematica, topologia. Si procede quindi alla redazione della reportistica con riferimento ai risultati dei controlli precedenti e riportando la documentazione delle fonti utilizzate e delle relative datazioni.

I controlli permettono di individuare anche i falsi errori e devono essere reiterati più volte a monte e a valle dell'editing sui dati.

Per i vari controlli di qualità sono state definite delle soglie di conformità alle specifiche (in percentuale) e quindi di accettabilità minima, che dovrebbero essere omogenee per i dati di tutte le regioni. Le soglie sono da documentare nella reportistica e possono essere anche molto basse ma comunque realistiche, in modo da permettere di valutare l'utilizzabilità del DB da parte di utenti potenziali.

Importante è inoltre l'aspetto della datazione dei dati utilizzati: occorre documentare la datazione dei dati derivati dalle regioni o da altre fonti e, per quanto riguarda i dati inseriti da noi, l'anno (o il triennio) delle ortofoto AGEA utilizzate. Queste informazioni sono presenti nel metadato operativo. Per i dati regionali occorre considerare che spesso sono riferiti a range temporali estesi (anche un decennio).

I dati numerici raccolti dai controlli possono essere successivamente elaborati per conteggi statistici e rappresentazioni grafiche.

4.1 Controllo della struttura

Questo tipo di controllo viene fatto in automatico mediante una procedura sviluppata internamente. Si effettuano inoltre e si analizzano i conteggi totali degli oggetti presenti in ogni classe per individuare criticità di informazione. Si rappresentano i risultati in tabelle da riportare nella reportistica finale.

4.2 Controlli di accuratezza tematica e di completezza

Questi controlli costituiscono una fase interattiva di lavoro, solo in parte automatizzabile. Occorre scegliere delle aree campione corrispondenti in totale a circa il 2% del territorio analizzato (cioè quello regionale, ad esempio scegliendo 3-4 superfici comunali significative).

Per la valutazione della completezza sono trascritti i casi di eccesso ed omissione per una selezione di oggetti considerati prioritari prendendo a riferimento le ortofoto aggiornate dell'AGEA come termine di confronto con il mondo reale.
Si utilizzano delle tabelle predisposte ad hoc per riportare le osservazioni ed effettuare i conteggi espressi in percentuale tra il numero degli elementi mancanti o in eccesso ed il numero di oggetti totali presenti sul terreno.
Per la valutazione dell'accuratezza tematica, di classifica e di congruenza logica degli oggetti, si utilizzano una serie di controlli sviluppati in ambiente ESRI (Data reviewer) che forniscono risultati in tabelle generate automaticamente. Le procedure controllano la classificazione dei oggetti contenuti all'interno di determinate pertinenze, ad esempio gli edifici classificati come sedi di ospedali nelle pertinenze ospedaliere. Sono effettuati anche controlli incrociati di attribuzione all'interno di classi.

4.3 Controllo di accuratezza posizionale

Si basa sul controllo del posizionamento di una nuvola di punti scelti su tutti i tipi geometrici da confrontare con i punti trigonometrici dell'IGM ritenuti idonei allo scopo. Vengono effettuate misure interattive sia su oggetti editati internamente sia per oggetti acquisiti esternamente considerando valori di accuratezza dichiarati dall'ente produttore della fonte utilizzata.
La valutazione è ottenuta in base allo scostamento delle componenti direzionali (est e nord) tra punto del DBSN e punto trigonometrico.

4.4 Controllo topologico

I controlli sono da realizzare con procedure automatiche e da limitare solo ai vincoli essenziali previsti dalla specifica di contenuto. Riguardano sostanzialmente:
- relazione spaziale tra oggetti di classi diverse con operatori overshoot/undershoot o intersect
- geometrie invalide
- duplicazioni di geometrie
- connessioni tra linee e nodi della viabilità (escludendo el_vms) e dell'idrografia (verificando anche i tratti fittizi e virtuali);
- inclusione di oggetti (edifici, edifici minori, manufatti) nelle pertinenze relative;
- verifiche di sovrapposizioni anomale.

Anche per questi controlli sono utilizzate procedure automatiche sviluppate in ambiente ESRI (data reviewer o model builder) che comprendono gli aspetti di validazione geometrica dei singoli oggetti e di relazione spaziale tra gli oggetti.

L'esempio di tabella riportata (Fig.8) mostra un raggruppamento per tipologia di controlli (indicati nella prima colonna come *Check Type*), e a seguire: i codici alfanumerici delle classi/attributi interessati (*Check Title*), il numero degli oggetti validati (*Records Validated*, cioè tutti quelli presenti nel database della regione Calabria), il numero degli errori risultanti (*Total Results*) e la percentuale della correttezza dei dati riscontrata (*Total Percent Accuracy*), valore nullo nel caso di assenza di oggetti.

Check Type	Check Title	Records Validated	Total Results	Total Percent Accuracy
Cutbacks Check				
	el_fer_cutbacks	6514	0	100,00000%
	el_vms_cutbacks	426476	37	99,99132%
	tr_ele_cutbacks	14103	15	99,89364%
	tr_str_cutbacks	440391	1	99,99977%
Duplicate Geometry Check				
	ab_cda_duplicati	21887	0	100,00000%
	ar_mar_duplicati	118	0	100,00000%
	att_sp_duplicati	6515	0	100,00000%
	condot_duplicati	168	0	100,00000%
	cs_mar_duplicati	3587	0	100,00000%
	cv_aers_duplicati	524	0	100,00000%
	cv_dis_duplicati	88	0	100,00000%
	diga_duplicati	20	0	100,00000%
	edi_min duplicati	350790	66	99,98119%
	edifc_duplicati	740770	0	100,00000%
	el_fer_duplicati	6514	0	100,00000%
	el_idr_duplicati	468350	0	100,00000%
	el_vms_duplicati	426476	310	99,92731%
	invaso_duplicati	1794	0	100,00000%
	mn_ind_duplicati	33050	0	100,00000%

Fig. 8. Esempio di tabella dei risultati del controllo.

5 Cartografia derivata dal DBSN

Fig. 9. Vestizione cartografica del DBSN.

IGM, storicamente, ha prodotto una *base map* tipicamente alla scala 1:25000, e successivamente ha ottenuto le carte alle scale più piccole derivando progressivamente dalla serie immediatamente inferiore. Questo anche in epoca di raccolta di dati in forma digitale.
Molto più ergonomico risulta invece avere i dati raccolti in un database con una risoluzione informativa 1:10000 per poi essere in grado, con diversi strumenti di generalizzazione, di derivare i dati alle varie scale di rappresentazione.
Nella nuova veste cartografica alla scala 1:25000 (Fig. 9), il taglio dell'elemento è piano quindi coincide con valori interi del reticolo cartografico e la sua ampiezza è di 18 Km in est e 14 Km in nord (252Km quadrati). Il reticolato è il prodotto dalla trasformazione cartografica Trasversa Mercatore di meridiani e paralleli, appartenenti al Sistema di Riferimento ETRF2008, applicata ad un unico fuso con origine in corrispondenza del meridiano centrale dei 12° (Fuso Italia).
il layout cartografico è la visualizzazione dei dati contenuti nel DBSN, almeno di quelli rappresentabili cartograficamente. Considerando che nel database le informazioni devono essere raccolte costantemente perché, in special modo alcuni tematismi, soggette a continuo aggiornamento, le operazioni di vestizione cartografica devono essere automatiche. In effetti sono applicate alcune procedure di generalizzazione cartografica necessarie per la leggibilità della carta, facilitate dal fatto che il passaggio da una risoluzione informativa alla scala nominale 1:10000 a quella ad una scala 1:25000 non necessita di operazioni onerose e complicate.
Come da legenda sono rappresentati tutti gli elementi considerati di primaria importanza e significatività evidenziati con diverso colore e simbologia per una lettura e interpretazione immediata del territorio.
Le carte sono prodotte in vari formati digitali ed in forma analogica.

Conclusioni

Il DBSN, per quanto siano condotti interventi e controlli per il miglioramento e l'aggiornamento continuo delle informazioni, secondo quanto sopra descritto, mantiene criticità ed errori retaggio dei dati originari utilizzati sia per le incompletezze che per la datazione di acquisizione. Tali criticità rimangono dato che il lavoro del DBSN si è concentrato su classi prioritarie e non su tutto il contenuto e anche per l'elevata numerosità di elementi di un database nazionale e l'ampia varietà di casistiche reali non trattabili sempre allo stesso modo.

Nella progettazione futura delle attività si cercherà di dare sempre più spazio ad applicazioni automatiche e sistematiche riducendo al minimo interventi manuali e interattivi puntuali, questi ultimi tra l'altro legati ad applicazioni e interpretazioni soggettive.

Un manuale operativo è stato redatto per guidare gli operatori ad elaborare i dati secondo stesse modalità, in base ad esempi grafici e raccomandazioni, ciò nonostante gli interventi vengono eseguiti con diversa precisione e osservanza dei criteri. Il manuale sarà comunque ulteriormente arricchito considerando tutte le particolarità e le eccezioni di oggetti che si presentano nei dati delle varie regioni.

Tuttavia l'enorme sforzo compiuto internamente dal personale tecnico non porta ad un database nazionale di qualità elevata su tutto il contenuto per cui è auspicabile che si prevedano contributi esterni direttamente dai produttori regionali ma anche dagli altri enti considerati fonte di dati primari.

Una collaborazione diretta, anche attraverso le nuove regole di interoperabilità stabilite dalla normativa recente, permetterebbe di raccogliere dati aggiornati e nuovi dati da ulteriori produttori. In particolare il Decreto legislativo 24 gennaio 2006, n.36 e s.m.i. (art.12 bis) relativo all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico, stabilisce il ruolo primario all'IGM per la raccolta, produzione, aggiornamento, riproduzione e distribuzione dei dati geospaziali definiti di elevato valore dalla Direttiva europea 1024/2019. L'entrata in vigore del decreto e del relativo documento di Linee guida consentirà all'Istituto di acquisire direttamente i dati con modalità informatiche (procedure API o download dei dati).

Da subito comunque possono essere stabiliti criteri e metodi di scambio dati in accordo con le regioni in particolare, nonché per dati di altri ministeri al fine di arricchire e ottimizzare il database sulla base delle esigenze comuni e nell'ambito dei propri compiti istituzionali.

Riferimenti bibliografici

1. Dulli Susi; Furini Sara; Peron Edmondo. Data Mining, Springer Verlag, 2009
2. Decreto 10 novembre 2011 Regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici. (Gazzetta Ufficiale n. 48 del 27/02/2012 - Supplemento ordinario n. 37). Allegato 1 - Catalogo dei dati territoriali – Specifiche di contenuto per i DataBase Geotopografici. Allegato 2 - Il modello GeoUML – Regole di interpretazione delle specifiche di contenuto per i DataBase Geotopografici