

## IL PROCESSO DI VALIDAZIONE DI UN DATABASE 3D MULTISORGENTE E MULTIRISOLUZIONE PER LE INFRASTRUTTURE FERROVIARIE

### THE VALIDATION PROCESS OF A 3D MULTISOURCE/MULTIRESOLUTION MODEL FOR RAILWAY INFRASTRUCTURES

M. Corongiu <sup>a</sup>, G. Tucci <sup>a</sup>, F. Flamigni <sup>b</sup>, A. Comparini <sup>b</sup>, F. Panighini <sup>a</sup>, E.I. Parisi <sup>a</sup>, L. Arcidiaco <sup>c</sup>

<sup>a</sup> Laboratorio SCHEMA (Survey, Cultural Heritage, Monitoring, Analysis) - Dipartimento di Ingegneria Civile ed Ambientale (DICEA), Università di Firenze, Via di S. Marta 3, 50139, Firenze (Italy)  
(grazia.tucci, manuela.corongiu, francesca.panighini, ericaisabella.parisi)@unifi.it

<sup>b</sup> Geoin srl, Via Panciatichi 110, 50127 Firenze (Italy) – (f.flamigni, a.comparini)@geoin.it

<sup>c</sup> Consiglio Nazionale delle Ricerche, Istituto di Bioeconomia (CNR IBE). Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), (Italy) - lorenzo.arcidiaco@ibe.cnr.it

**PAROLE CHIAVE:** qualità dati geografici, specifiche, infrastruttura ferroviaria, procedure di valutazione, certificazione, database spaziali multisorgente, BIG data

**KEY WORDS:** geographic data quality, data product specification, railway infrastructure data, evaluation procedures, certification, multisource spatial databases, BIG data

**Questo articolo è il sunto tradotto in italiano della versione sottomessa allo Special Issue “The Impact of Innovative and Emerging Technologies on the Surveying Activities” della rivista APPLIED GEOMATICS collegato al convegno SIFET 2018.**

#### RIASSUNTO

L'evoluzione tecnologica che segna il passaggio dalla cartografia tradizionale ai database topografici è spesso accompagnata, a livello di rilievo, dall'adozione di diverse fonti e strumenti per l'acquisizione di dati: telerilevamento, droni, laser scanner, ecc. Questo approccio ha però un impatto significativo sulla validazione dei dataset geografici: ogni singola fase deve tenere conto dello specifico sensore utilizzato e delle sue caratteristiche tecniche, ma anche delle congruenze tra i diversi passaggi che contribuiscono alla creazione di un database. Pertanto, la specifica del prodotto finale è solo il primo passo verso la certificazione della qualità del dataset, cui andrà aggiunta anche la correttezza dei processi intermedi eseguiti. In tale contesto, dove la standardizzazione dei processi di acquisizione non ha raggiunto un livello di stabilità consolidata, il presente contributo fa il punto sulle questioni critiche incontrate durante un processo di certificazione di qualità dei prodotti per l'infrastruttura geografica nel contesto ferroviario. Il processo di validazione è stato condotto per gradi. Per i prodotti aerofotogrammetrici tradizionali, la validazione ha seguito una prassi ormai consolidata mentre per quelli più innovativi acquisiti da treno con MMS (Mobile Mapping Systems) ha seguito una profilazione degli standard ISO sulla qualità. Infine, per il DBGT (database geotopografico) sono state combinate procedure massive (informatiche) con quelle a campione per le accuratèzze tematiche/metriche delle singole istanze. La notevole varietà e la quantità di dati forniti nella logica BIG Data, l'armonizzazione tra i diversi processi di valutazione, la tempistica serrata per la convalida dei vari lotti, sono stati ulteriori fattori caratterizzanti dei procedimenti attuati.

#### ABSTRACT

The technological evolution that marks the passage from traditional cartographies towards topographic databases is often characterised, at a surveying level, by the availability of different sources and tools for acquiring data: remote sensing, drones, laser scanner, etc.

These multiple sources have a significant impact during the dataset evaluation phases. For each specific sensor used, its technical characteristics have been taken into account. Therefore, data product specification is the first step towards understanding which spatial database quality and requirements must be satisfied. While sources and tools have yet to be established in the processing and methodologies, this article try to focus on critical issues encountered during the validation process of a geographical infrastructure in the railway context. The validation process has been carried out by a step-by-step approach. Basically, a consolidated validation methodology has been adopted for traditional products also carried out by new sensors, while a comparison with ISO standard specifications has been followed for innovative surveys such as MMS (Mobile Mapping Systems). Finally, for the GeoTopographic DataBase (DBGT) both massive (informatic procedures) and traditional thematic evaluation of accuracy have been combined. Therefore, the adherence with standards have been referred to consider both the quality of data and the conformity to data product specifications.

The considerable variety and the amount of the provided data in term of BIG Data, the harmonization between different evaluation processes, the need to validate with short time, the lack of on-the-field surveys, have been hardly considered as additional requirements of quality certification of the adopted procedures.

## 1. INTRODUZIONE

Negli ultimi anni, i dati spaziali sono stati implementati combinando e armonizzando diversi dataset a loro volta acquisiti tramite diversi sistemi e sensori: rilievi aerei, rilievi terrestri, droni, LiDAR, ecc. Di conseguenza, il processo di validazione multirisoluzione e multiaccuratezza ha assunto connotazioni complesse e non consolidate rispetto ai criteri di certificazione della qualità di ogni singolo dataset, oltre che alla congruenza di tutte le informazioni in un database spaziale integrato. Il modello di gestione che ne deriva, non può che afferrare ad un approccio da BIG data per l'onerosità e la complessità dei dati trattati.

Questa notevole quantità di dati raccolti da diverse fonti ha aperto a una più ampia gamma di possibilità e approfondimenti per futuri miglioramenti tecnologici che copriranno nuove aree di applicazione (Thaduri et al. 2015).

Insieme alle criticità da BIG data, quindi, devono essere presi in considerazione due aspetti: le diverse fasi di fornitura e l'evoluzione della cartografia in banche dati topografiche, di seguito spiegate.

### 1.1 Le diverse fasi di fornitura di dati spaziali multidimensionali

Nel contesto di un processo di produzione di mappe cartografiche, un aumento significativo della disponibilità di diversi strumenti e sensori porta a una maggiore complessità delle strutture dati. In effetti, il classico processo di acquisizione è stato, in passato, principalmente ottenuto per via aerofotogrammetrica, finalizzato alla creazione di una mappa in un momento specifico per un determinato territorio. In questo caso, i processi, i tempi e le metodologie per la validazione dei dataset possono essere considerati stabili e standardizzati. Lo stesso non si può dire per un processo dove l'acquisizione avvenga con modalità multisorgente e multirisoluzione. In tale circostanza, il processo di validazione di ogni singola fase dovrebbe essere eseguito in conformità alla tecnologia adottata e successivamente integrato in un'infrastruttura geografica complessa.

Allora il prodotto cartografico finale sarà il risultato di un'elaborazione combinata di diverse fonti di dati. Il processo di implementazione sarà quasi sempre costituito da una sequenza di passaggi in cui il risultato di una fase sarà il punto di partenza di quella successiva. Le modalità di validazione di ciascun prodotto intermedio dipenderanno dalla gestione del processo da eseguire.

### 1.2 Cartografia tecnica vs database spaziali

In un contesto tradizionale, la cartografia di riferimento è redatta da un'entità istituzionale competente. In Italia, una CTR (Carta Tecnica Regionale) è realizzata da Enti Locali e rappresenta una base cartografica a grande scala (1:2.000, 1:5.000, 1:10.000). Per gestire grandi infrastrutture di trasporto a sviluppo longitudinale su territorio come strade, ferrovie, ecc., si necessita di dati spaziali sufficientemente precisi ma che coprano ampi territori, a cavallo di diverse Regioni. Se si vuole inquadrare congruentemente il dataset di una infrastruttura di trasporto con il territorio circostante occorre farsi carico dell'armonizzazione delle diverse CTR, tra loro, e con l'infrastruttura di trasporto considerata.

Inoltre, recentemente, il prodotto cartografico si è evoluto, passando in pochi decenni da una mappa di riferimento di base a complessi dataset composti da geometrie 3D, vincoli topologici, relazioni con i database, immagini, modelli

numerici, ecc. Spesso ogni contenuto deriva dall'interazione tra diversi sensori integrati con procedure implicite e complesse di cui non rimane traccia nel prodotto finale. Pertanto, anche l'attività di certificazione della qualità è diventata complessa e non è in grado di ignorare l'aspetto intermedio, soprattutto se ha contribuito al prodotto finale.

Al giorno d'oggi, dove la cartografia è diventata un database spaziale, l'esigenza di garantire un corretto livello di interoperabilità richiede un database topografico di riferimento che sia sufficientemente robusto da resistere all'evoluzione tecnologica e abbastanza flessibile da ricevere gli aggiornamenti continui nel tempo. Tuttavia, con riferimento alla specifica per i DBG (database geotopografici) (DM, 2012), esse si riferiscono a specifiche di contenuto in termini di modello concettuale, quindi qualsiasi Ente Locale può implementarle secondo le proprie esigenze a livello fisico e conformemente alle specifiche necessità di accuratezza spaziale e temporale.

#### 1.2.1 Verso il 3D city model

Il 3D city model considera una rappresentazione di un ambiente urbano con una geometria tridimensionale di tutti gli oggetti e strutture presenti, in cui gli edifici sono le caratteristiche più importanti. La definizione di requisiti attraverso domini spaziali 3D tematici è stata descritta in studi recenti (Biljecki et al., 2015). Tuttavia, in passato, il 3D city model è stato principalmente adottato per scopi di visualizzazione, ma al giorno d'oggi, un numero crescente di applicazioni, ha necessitato di ulteriori informazioni rispetto a quelle disponibili per le sole rappresentazioni ancorché standardizzate (Kolbe, 2009). Solo per citare alcuni esempi: simulazioni e analisi dei trend di cambiamenti del suolo, pianificazione e programmazione urbana, gestione delle catastrofi, sicurezza nazionale sono solo alcuni esempi di applicazioni che necessitano di analisi anche complesse sul territorio basate su banche dati opportunamente strutturate.

#### 1.2.2 Standard sul modello di rete ferroviaria

Il riferimento standard per la modellizzazione delle informazioni spaziali tematiche è rappresentato dalla Direttiva INSPIRE (INSP., 2007). Uno dei principali riferimenti standard nel contesto delle ferrovie è rappresentato dalla specifica sulle reti di trasporto (INSP.D2.8.I.7, 2014) che definisce il modello di rete di trasporto. In questo standard, i dati si riferiscono ad oggetti topografici correlati al trasporto stradale, ferroviario, marittimo e aereo e fornisce un approccio coerente ai modelli spaziali di rappresentazione (ingombro spaziale e struttura a grafo della rete).

Tutti i temi di dati spaziali devono seguire le indicazioni della Direttiva INSPIRE in termini di modello concettuale GCM (Generic Conceptual Model) (INSP.D2.10.1, 2013) e GNM (Generic Network Model) (INSP.D2.10.1, 2013). Il GNM comprende la definizione di: i) meccanismi di connessione di rete per stabilire la connettività transfrontaliera o intermodale, ii) riferimento a oggetti per supportare il riutilizzo delle informazioni e iii) riferimento lineare per supportare e collegare le diverse proprietà di trasporto agli elementi di trasporto. A loro volta si basano su diverse serie ISO di standard di informazione geografica per fornire le basi di aspetti specifici dell'interoperabilità in particolare con lo standard sui sistemi di trasporto intelligenti come l'LBS (Location Based Services) (ISO19148, 2012).

Per quanto riguarda il modello topologico geometrico delle reti di trasporto (INSP.D2.8.I.7, 2014), le specifiche dei dati definiscono uno schema concettuale a grafo con eventi proiettati come posizioni singole su questo e memorizzandone la

progressiva metrica. Per tale motivo il posizionamento sulle strutture a grafo avviene adottando il sistema di riferimento lineare (Linear Reference System) basato sul concetto di ascissa curvilinea. Tale modello è adatto per trasporti, servizi pubblici, servizi basati sulla posizione e altre applicazioni che definiscono posizioni relative ad oggetti lineari. La specifica dei dati include tre tipi di geometria: (a) (area d'interesse), (b) linea della mezzeria come proiezione di un dato evento lungo l'elemento di grafo e (c) punti degli oggetti. I tipi (a) e (b) possono essere rappresentazioni alternative degli stessi fenomeni del mondo reale a cui gli utenti possono associare le proprie informazioni (oggetti). Il tipo (c) rappresenta sia i nodi della rete di trasporto che i punti d'interesse da proiettare su grafo. Il tipo di rappresentazione spaziale di base è un vettore 2D.

La topologia viene gestita nella specifica dei dati in modo implicito piuttosto che esplicito, principalmente al fine di mantenere il modello il più semplice possibile poiché si prevede che la maggior parte delle applicazioni utilizzerà i dati di rete in un ambiente topologico. Di conseguenza esiste il prerequisito per la "topologia implicita", in cui i dati forniti devono essere sufficientemente puliti e in grado di costruire una topologia automatizzata all'interno dell'applicazione dell'utente. Questo concetto è inquadrato da requisiti specifici, comprese le informazioni sulla qualità dei dati. In generale, anche se i modelli di rete sono semanticamente ricchi e complessi, rappresentano solo la topologia 2D della rete oltre alle informazioni logiche sulla connettività di rete. Una rappresentazione della geometria 3D potrebbe essere associata all'uso di funzionalità multipatch, tuttavia queste non sarebbero accoppiate alla modellazione di rete, ma solo a una visualizzazione 3D pura. Ciascuno dei modelli di dati di dominio rappresenta un'astrazione 2D della rispettiva rete ferroviaria nel mondo reale. Ogni modello di rete ferroviaria è specificatamente definito, non facilmente utilizzato per differenti tipi di rete ferroviaria, nel contempo nello standard non viene fornito alcun modello/database comune che integri modelli diversi. Tuttavia, il GNM consente la tracciabilità di un network o altri tipi di analisi dello stesso (Adolphi et al., 2013).

## 2. OGGETTO: LA VALIDAZIONE DI DATASET GEOTOPOGRAFICI 3D E MULTISORGENTE

Questo articolo è focalizzato sulla descrizione degli aspetti critici incontrati durante le fasi di validazione dei dati spaziali nel contesto delle ferrovie, prendendo in considerazione sia metodiche tradizionali che innovative, integrandole, al fine di certificare la qualità dei dataset multisorgente e dei processi attuati per acquisirli. Il processo di validazione è stato condotto con un approccio graduale di validazione step-by-step delle singole fasi elementari. Fondamentalmente, è stata adottata una metodologia di verifica ormai consolidata per i prodotti di acquisizione aerofotogrammetrica tradizionale, mentre per i prodotti più innovativi e recenti, quali ad esempio le nuvole di punti acquisite da laser scanner montato su treno mediante sistemi MMS (Mobile Mapping Systems), si è fatto riferimento agli standard ISO sulla qualità dei dati geografici. Infine, per il DBGT sono state combinate procedure massive (informatiche) proprie del database con quelle a campione per le accuratezze tematiche/metriche delle singole istanze.

Per garantire la conformità tra i diversi prodotti di fornitura, la valutazione della qualità è stata effettuata innanzitutto al termine ogni fase del processo produttivo, diventando questa propedeutica per la validazione di quella successiva. Inoltre, è necessario un ripensamento del processo di validazione nel

contesto di banche dati spaziali per tener conto della sorgente (e quindi dell'accuratezza) di acquisizione di una certa informazione, oltre che delle congruenze tra classi di oggetti nel loro insieme. Le procedure dovrebbero mirare a convalidare la conformità del contenuto e della struttura dei dati con il modello spaziale indicato nelle specifiche di contenuto, compresa la geometria e i vincoli topologici (Carrion et al., 2008). Infatti, la principale differenza tra un prodotto tradizionale della cartografia ed i contenuti di una più complessa SDI (Spatial Data Infrastructure) è la dipendenza di ogni fase da quella successiva in termini di prodotto di qualità e in generale, come coerenza tra le diverse parti del dataset. In questo contesto, la certificazione di qualità non si è fermata alla verifica di conformità rispetto alla specifica del prodotto, ma si è spinta a controlli di tipo analitico di congruenza tematica e spaziale sia dei singoli oggetti territoriali che in funzione del contesto nel quale si inseriscono.

Con riferimento allo standard sulle specifiche di prodotto (ISO19131, 2018), queste definiscono una descrizione dettagliata di uno o più dataset, corredati da informazioni aggiuntive che ne consentano la creazione, la fornitura e l'utilizzo da parte di un altro soggetto. In sintesi, lo standard definisce come formalizzare una descrizione tecnica precisa e formale del prodotto geografico che si vuole ottenere. Lo scopo di questo standard internazionale è quello di indicare come fornire i requisiti sul contenuto delle specifiche del prodotto in conformità con altri standard esistenti per le informazioni geografiche.

Quindi, conformemente a quanto indicato in questo standard, le specifiche per i DBGT (DM, 2012) sono state integrate con le specifiche tecniche di fornitura e le specifiche amministrative oltre che integrate con dati specifici del caso in esame (ambito ferroviario) dato che, come dichiarato nelle specifiche per i DBGT (DM, 2012), i dati territoriali ivi descritti costituiscono solamente un primo nucleo informativo da integrare con altri database tematici.

### 2.1 Specifiche di prodotto dei dati geografici e standard di qualità

La descrizione della qualità dei dati geografici facilita il confronto e la selezione di dataset su misura, per esigenze e requisiti specifici dell'applicazione. Descrizioni complete sulla qualità dei dati incoraggiano la condivisione, l'interscambio e l'uso appropriato dei dataset. Le informazioni sulla qualità dei dati geografici consentono al produttore di dati di valutare in che misura un dataset soddisfi i criteri stabiliti nelle specifiche di prodotto e aiuta gli utenti a valutare la capacità di un prodotto verso i requisiti di una particolare applicazione d'interesse. Ai fini della valutazione della qualità, sono state utilizzate le coerenze delle procedure definite nello standard sulla qualità dei dati geografici (ISO9157, 2013). Le diverse fasi e livelli di azioni necessarie alla valutazione di qualità sono state le seguenti:

- Risultati finali;
- Validazione;
- Accessibilità;
- Aggiornamento;

Considerando l'azione 1, un aspetto critico ha riguardato le fasi di assemblaggio dei risultati derivanti da ogni controllo specifico con le congruenze di aggancio tra lotti omogenei, oltre alla messa a sistema di processi dove la validazione di ogni step risultasse propedeutico all'avvio delle validazioni agli step successivi. In generale, i pesi da attribuire a ciascuna fase, le modalità di armonizzazione di diversi dataset multisorgente, la

certificazione di qualità complessiva del database, sono solo alcuni degli aspetti peculiari cui si è dovuto far fronte nella gestione delle fasi di certificazione di qualità.

Con riferimento allo standard sulle specifiche di prodotto dei dati geografici (ISO19131, 2018) si è dovuto tener conto di quanto richiesto nel caso studio in merito alle tempistiche ed alle congruenze nelle forniture dei vari lotti, in quanto esse hanno un impatto sul processo di valutazione sia della qualità in ciascun lotto che di una certificazione generale dei risultati.

Con riferimento allo standard sulla qualità dei dati geografici (ISO19157, 2013) occorre considerare diversi “elementi” e “sotto-elementi”. Per descriverli, vengono utilizzate le “unità di qualità dei dati”, ovvero la combinazione di “scopo” ed “elementi di qualità”. Gli “elementi di qualità dei dati” sono componenti che descrivono determinati aspetti della qualità dei dati geografici, organizzati in diverse “categorie”.

Le “categorie” considerate nella valutazione della qualità per il caso studio sono:

- Completezza (C1): definita come presenza e assenza di caratteristiche, attributi e relazioni. Consiste in:
  - commissione (C1\_1): dati in eccesso presenti in un set di dati;
  - omissione (C1\_2): dati assenti da un set di dati;
- Coerenza logica (C2): definita come il grado di aderenza alle regole logiche della struttura, dell'attribuzione e delle relazioni logiche (la struttura dei dati può essere concettuale, logica o fisica). Se queste regole logiche sono documentate altrove (ad esempio, in una specifica del prodotto dati), è necessario farne esplicito riferimento. Consiste in:
  - coerenza concettuale (C2\_1): aderenza alle regole dello schema concettuale;
  - coerenza del dominio (C2\_3): aderenza dei valori ai domini dei valori;
  - coerenza del formato (C2\_4): grado in cui i dati vengono archiviati conformemente alla struttura fisica del set di dati;
  - coerenza topologica (C2\_5): correttezza delle caratteristiche topologiche esplicitamente codificate di un set di dati;
- Precisione posizionale (C3): definita come accuratezza della posizione di un dato all'interno di un sistema di riferimento spaziale. Consiste in:
  - precisione assoluta o esterna (C3\_1): vicinanza dei valori di coordinate assegnati a valori accettati come veri;
  - accuratezza relativa o interna (C3\_2): vicinanza delle posizioni relative dei dati alle rispettive posizioni accettate come vere;
  - precisione della posizione dei dati con griglia (C3\_4): vicinanza dei valori di posizione spaziale dei dati grigliati su valori accettati come veri;
- Accuratezza tematica (C4): è definita come l'accuratezza degli attributi quantitativi e la correttezza degli attributi non quantitativi rispetto alle classificazioni delle istanze e delle loro relazioni. Consiste in:
  - correttezza della classificazione (C4\_1): confronto delle classi o dei loro attributi assegnati a un universo di discorso (ad es. Verità di base o dati di riferimento);
  - correttezza dell'attributo non quantitativo (C4\_2): misura della correttezza o meno di un attributo non quantitativo;
  - accuratezza dell'attributo quantitativo (C4\_3): vicinanza del valore di un attributo quantitativo a un valore accettato come o noto per essere vero;

• Qualità temporale (C5): è definita come la qualità degli attributi e delle relazioni temporali del dataset. Consiste in:

- accuratezza di una misurazione del tempo (C5\_1): vicinanza delle misurazioni del tempo riportate a valori accettati o conosciuti come veri;
- coerenza temporale (C5\_2): correttezza dell'ordine degli eventi;
- validità temporale (C5\_3): validità dei dati rispetto al tempo;

Infine, abbiamo anche considerato la “categoria”:

• Misure di qualità dei dati (D1): un elemento di qualità dei dati dovrebbe riferirsi a una sola misura di qualità del dato, scelta come riferimento univocamente e completamente descritta tramite un identificatore ed una descrizione associata.

### 3. IL CASO STUDIO: INFRASTRUTTURE DELLA RETE FERROVIARIA

L'infrastruttura ferroviaria di cui parleremo in questo articolo si riferisce ai dataset forniti per il progetto italiano MUIF (Modello Unico dell'infrastruttura Fisica), sull'infrastruttura di RFI (Rete Ferroviaria Italiana). Il laboratorio SCHEMA (laboratorio congiunto tra Università degli Studi di Firenze e L'Istituto Geografico Militare), in tale contesto, è responsabile della certificazione di qualità dei prodotti cartografici.

A livello generale, il progetto MUIF ha lo scopo di gestire tutte le informazioni richieste dall'infrastruttura ferroviaria, in un unico modello, in termini di acquisizione e accessibilità dei dati, strutturati congruentemente per essere serviti attraverso le applicazioni WebGIS-WebService (Corongiu et al., 2018).

#### 3.1 I dataset multisorgente

I dataset del MUIF utilizzati come caso studio si riferiscono ai seguenti prodotti:

- ortofoto (Ground Sample Distance - GSD=0.10 m);
- DTM (Digital Terrain Model) - DSM (Digital Surface Model) (1p.to/m<sup>2</sup>);
- DBGT integrato e multiprecisione, alla scala 1:1.000 nelle stazioni ferroviarie, alla scala 1:2.000 in aree di rispetto di 120m/500m lungo la tratta ferroviaria, Figura 1.

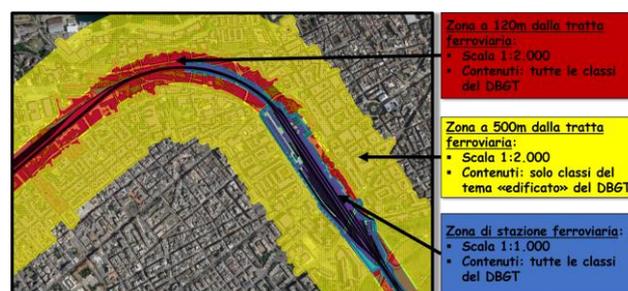


Figura 1 - database multirisoluzione

I dataset sono stati acquisiti da:

- Fotogrammetria aerea (GSD=0.08 m);
- Rilievo LiDAR da aereo (4 punti/m<sup>2</sup>);
- Rilievo MMS da treno con i seguenti prodotti: immagini side-by-side, immagini a 360° (sei scatti sincroni), nuvole di punti di telemetria con l'attribuzione del colore RGB (Red Green Yellow) da immagini a 360° (Fig. 2).



Figura 2 - Nuvole di punti da laser scanner telemetrico montato su treno

Nel contesto ferroviario, i dataset d'interesse sono caratterizzati da un significativo livello di dettaglio sia nella dimensione spaziale che in quella temporale, integrati congruentemente nel contesto topografico nel quale si inseriscono.

Per quanto riguarda le informazioni ferroviarie di dettaglio, vi sono alcune specificità da tenere in conto nel processo di validazione. Le recenti politiche europee stabiliscono che i gestori delle infrastrutture ferroviarie debbano concentrarsi sulla riduzione dei costi operativi ma, allo stesso tempo, aumentare le prestazioni sia delle attività finanziarie che della sicurezza. Per perseguire tali obiettivi, dal punto di vista tecnologico, uno dei componenti principali del sistema ferroviario è costituito da un CMMS (Computerized Maintenance Management System) che di fatto, implementando una LAM (Linear Asset Management), gestisce il posizionamento e l'anagrafica di ASSET ferroviari in un sistema "a grafo". Conseguentemente, l'utilizzo della segmentazione dinamica consente la proiezione di eventi lungo la linea di mezzieria del binario caratterizzandoli come attributi della rete ferroviaria. Il CMMS consente inoltre la descrizione delle relazioni lineari che indicano intersezioni e incroci paralleli o di pendenza di attività lineari.

L'implementazione di un sistema di monitoraggio delle risorse ferroviarie basato su un modello di dati semantici offre vantaggiose possibilità di integrazione, adattabilità e compatibilità dei dati rispetto agli approcci tradizionali. La mancanza di incentivi all'interoperabilità negli ultimi anni ha generato uno scenario non coordinato in cui diverse compagnie spesso gestivano rotte concorrenti tra le stesse destinazioni. In tale contesto, risulta evidente come l'impatto in termini di interoperabilità tra diversi modelli di dati sia il principale aspetto critico su cui concentrarsi per trovare una soluzione comune anche nella gestione dei dati spaziali ferroviari (Thaduri et al., 2015).

### 3.2 Le specifiche del prodotto dei dati per il DBGT

In termini di specifiche del prodotto dei dati geografici per il caso studio, le norme per i DBGT (DM, 2012) sono state profilate e armonizzate per consentirne l'implementazione

fisica. Pertanto, i contenuti sono stati organizzati in un unico database integrandolo ASSET ferroviari specifici. Esempio di ASSET sono; deviatoi, intersezioni, manufatti ferroviari, portali di alimentazione, ecc.). Questi sono stati integrati nel DBGT modellandoli coerentemente con gli altri oggetti di natura geotopografica (Fig. 3).



Figura 3 – oggetti semantici del DBGT integrato

Le risorse ASSET integrate nel DBGT mantengono il collegamento a ciascuna applicazione tematica specifica, tramite un identificatore esterno di connessione con tali archivi.

## 4. METODOLOGIE DEL PROCESSO DI VALIDAZIONE

### 4.1 La validazione nel contesto ferroviario

Per l'estensione della rete ferroviaria sul territorio nazionale, la fornitura è stata organizzata in lotti, e, per ogni lotto, suddivisa in funzione della fonte di rilevazione. Pertanto, i processi di validazione dei prodotti intermedi hanno tenuto conto della congruenza sia tra contenuti multisorgente che tra lotti. Per la prima questione di congruenza sono stati valutati diversi tipi di dataset (ortofoto vs. LiDAR, ecc.). Mentre per la seconda questione di congruenza sono stati considerati lotti adiacenti dello stesso tipo di dataset.

Di conseguenza, la certificazione di qualità dei dati spaziali non è stata ottenuta solo per semplice addizione del processo di validazione di ogni singolo tipo di dataset, ma andando a considerare anche tutte le dipendenze spazio/temporali, oltre alle congruenze richieste per dati multidimensionali, implementando regole, vincoli topologici e procedure in un unico DBGT. Questo approccio ha richiesto una definizione di un modello condiviso, tramite la realizzazione di una specifica di progettazione dei dati cartografici.

### 4.2 La validazione del rilievo fotogrammetrico aereo

Il rilievo aereo fotogrammetrico è stato profondamente rivoluzionato dall'avvento del digitale.

Le fotocamere a pellicola erano quasi tutte uguali per formato e prestazioni. A livello di lunghezze focali la scelta era ristretta a due tipologie: il grandangolo per la maggior parte degli impieghi e il teleobiettivo per le riprese finalizzate alla realizzazione di ortofoto. Le pellicole disponibili avevano una sensibilità standard per cui anche la risoluzione equivalente risultava costante. Tutto ciò comportava che l'intero processo di ripresa e di verifica potesse basarsi su set di parametri precostituiti. A una data scala cartografica corrispondeva una ben precisa scala fotogramma e questa poteva essere realizzata solo con una certa quota di sorvolo. Il processo di verifica e validazione risultava se vogliamo più semplice. Con il digitale ci si deve confrontare con molti sensori che variano per risoluzione e dimensione per la rapida evoluzione dei sensori optoelettronici. Le focali non sono più uno standard dovendosi sempre confrontare con la dimensione del sensore (sarebbe a

questo punto più corretto parlare di angoli di campo). La scala del fotogramma ha come equivalente nel digitale il GSD ovvero la dimensione della mattonella di territorio che corrisponde ad un pixel immagine che diventa quindi il parametro fondamentale per giudicare se un certo rilievo aereo è idoneo a rilevare cartografia con la scala (o livello di dettaglio) richiesta. Per misurare il GSD è necessario disporre del certificato di calibrazione della fotocamera, dei parametri di assetto dei singoli fotogrammi (oggi immediatamente disponibili perché forniti dal GNSS - Global Navigation Satellite System - e dall'IMU - Inertial Measurement Unit - dell'aereo) e di un modello numerico di dettaglio medio basso dell'area sorvolata (con una camera avente il pixel da 5µm ed una focale da 50mm una differenza di quota di 10 m produce una differenza del GSD di 1mm). Una volta ottenuta l'impronta al suolo di ogni fotogramma gli altri indicatori di qualità quali il ricoprimento dell'area oggetto del rilievo, l'adeguato ricoprimento longitudinale e trasversale dei fotogrammi possono essere analizzati in modo del tutto automatico. Ovviamente questi non sono i soli indicatori di qualità; occorre prendere in esame anche la qualità fotometrica delle immagini, l'eventuale presenza di elementi di disturbo quali nubi, incidenza delle ombre, per i quali è necessaria un'ispezione visiva anche se a campione.

Per gli aspetti metrici occorre considerare che da quando sono disponibili i sensori inerziali a bordo degli aerei il numero di GCP (Ground Control Point) per l'appoggio al suolo si è notevolmente ridotto.

Inoltre il costo esiguo del singolo fotogramma digitale (se confrontato col costo del fotogramma su pellicola) permette di acquisire fotogrammi con una notevole sovrapposizione, prerequisito fondamentale per poter impiegare tecniche SfM (Structure for Motion) per l'abbinamento automatico di punti omologhi (questo non ha effetti sulla precisione in quota in quanto la ridondanza nel numero di fotogrammi serve per poter "tracciare" in modo sicuro un certo particolare, in fase di misura il software provvederà a selezionare i soli fotogrammi con base adeguata). L'enorme numero di punti omologhi che è possibile utilizzare per la triangolazione aerea unitamente alla loro elevata confidenza ottenuta mediante procedure statistiche di eliminazione degli errori RANSAC (Random Sample Consensus) conferisce al blocco una grande rigidità, per cui la verifica potrà a sua volta basarsi su un esiguo numero di punti noti (check points) da rimisurare per stimare l'entità degli errori residui.

#### **4.3 La validazione di DTM/DSM acquisiti da rilievo LiDAR su aereo**

Con riferimento allo sviluppo territoriale di una infrastruttura di trasporto, l'orografia rappresenta uno dei dataset informativi fondamentali per comprendere connessioni, analisi di criticità, valutazione dei rischi, orientate alla pianificazione e gestione della relativa infrastruttura di dati territoriali. I sensori della tecnologia LiDAR creano strisciate grezze, mentre una fornitura generale è costituita da dataset organizzati in fogli di territorio. Il processo di suddivisione in tile è spesso uno strumento implicito all'interno del software in cui è difficile memorizzare informazioni sugli algoritmi utilizzati. Ciò rende il processo di certificazione della qualità di questo specifico aspetto piuttosto problematico. Sarebbe preferibile che la valutazione fosse effettuata considerando sia i prodotti finali che quelli intermedi (singole strisciate e nuvole di punti organizzate in fogli di fornitura), allo scopo di certificare sia la fornitura finale sia i processi seguiti.

Ad esempio, i file griglia sono il risultato di un processo di classificazione e interpolazione. L'interpolazione è efficace se i dati di origine sono sufficientemente densi considerando strisce grezze e nuvole di punti, per accertarsi che non esistano aree di sotto-campionatura. La correttezza della quota, inoltre, consente anche di verificare l'assenza di anomalie nei dati di origine. Quando si riscontrano anomalie solo nei file griglia, significa che si è verificato un errore durante il processo di interpolazione.

#### **4.4 La validazione del rilievo da treno**

Per quanto strettamente riguardante la sede ferroviaria, il dataset maggiormente significativo è rappresentato dalle nuvole di punti provenienti da laser scanner e immagini fotogrammetriche acquisite da treno. Per il caso studio il laser scanner è posizionato in coda al treno e rileva fino ad una distanza di circa 40 m dalla mezzera del binario, mentre le immagini stereoscopiche generano una nuvola di punti fino alla distanza di circa 25 m dalla mezzera del binario, con passo di acquisizione stabilito per il sistema di fotogrammetria di circa 2 m. Inoltre, nelle stazioni ferroviarie, dove la localizzazione degli asset ferroviari può essere esterna all'area coperta dal lidar da treno, vengono effettuati dei "rilievi aggiuntivi da terra" tramite laser scanner montati su zainetti, carrellini ferroviari, ecc. con caratteristiche simili alle precedenti. Ogni fase di integrazione nelle stazioni è comunque pianificata a priori.

Come accennato in precedenza, tutti i dataset forniti sono organizzati in lotti, come unità informative elementari omogenee, per garantire una buona gestione dei BIG data, consentendo così una valutazione della qualità dei risultati finali. Per ogni lotto, le nuvole di punti occupano circa 3-4 TB di spazio di archiviazione; poiché l'infrastruttura è divisa in 90 lotti, dobbiamo avere una dimensione dell'archivio di circa 300 TB. Con riferimento a quanto previsto nello standard (ISO19157, 2013), la valutazione della qualità è stata effettuata seguendo un approccio rigoroso sulla successione delle diverse fasi formali. Innanzitutto, il DQS (Data Quality Scope) è definito come gli elementi e le caratteristiche generali che identificano i dati su cui vengono condotti i controlli di qualità. Con riferimento al caso studio, gli elementi di base considerati sono stati:

- piano di acquisizione dei dati: definisce i dati da acquisire in termini di caratteristiche geografiche;
- GCP - punti di riferimento, misurati sul terreno, da utilizzare come parametro di confronto per valutare la precisione delle nuvole di punti;
- dati laser ferroviari: dati laser scanner acquisiti con MMS. Per ogni lotto esiste un solo tipo di dataset, ma può essere ottenuto da uno o più passaggi (corse) del treno sul quale è montato sia il laser scanner che il sistema fotogrammetrico, lungo la tratta ferroviaria. Conseguentemente gli output ottenuti saranno uno o più serie di nuvole di punti e una o più serie di immagini;
- nuvole di punti: costituito da una sequenza di dati laser, ordinati temporalmente, ogni acquisizione dà origine a una o più nuvole di punti. Per motivi di praticità nell'analisi, le nuvole di punti sono state a loro volta suddivise in sotto-nuvole (tronchi di nuvole di punti), lunghe circa 1 km sull'asse del binario. Tra tronchi adiacenti viene mantenuta un'area di sovrapposizione per consentirne la contiguità e corrispondenza;
- immagini 360: è una sequenza di immagini panoramiche acquisite simultaneamente ai dati laser, in modo che possa essere associato il loro valore RGB alla nuvola di punti laser.

Ogni immagine panoramica è geolocalizzata con punto 3D lungo l'asse del binario.

- dati fotogrammetrici da treno: si riferiscono a immagini stereoscopiche. Le immagini stereo sono ordinate temporaneamente, da ognuna di queste viene creata una nuvola di punti, grazie alle tecniche SfM. Ogni coppia stereo è geolocalizzata con punto 3D lungo l'asse del binario.

In linea con lo standard di qualità (ISO19157, 2013), gli elementi di qualità utilizzati sono:

- C1 - per verificare se tutti i file previsti dalle specifiche del prodotto siano stati forniti con la struttura appropriata;
- C2 - per verificare la congruenza sugli indici tra nuvole di punti ed immagini panoramiche, come anche la congruenza tra immagini stereoscopiche e nuvole di punti sintetiche;
- C2\_4 - per verificare immagini panoramiche e stereoscopiche per quanto riguarda formato e luminanza;
- C2\_5 - per verificare la contiguità delle immagini panoramiche, verificando anche che queste siano state acquisite nella frequenza definita per ogni corsa e che le nuvole di punti (sia laser che sintetiche) siano state correttamente colorate;
- C2\_5 - per verificare la copertura di immagini panoramiche e stereoscopiche rispetto al piano di acquisizione;
- C2\_5 - per verificare la copertura delle nuvole di punti laser considerando la loro impronta rispetto al piano di acquisizione;
- C2\_5 - per verificare la contiguità delle nuvole di punti laser, in modo che i tronchi di nuvole di punti siano collegati in modo continuo e prestando anche attenzione a quando si sovrappongono diverse corse;
- C3\_1 - per verificare la precisione delle nuvole di punti laser in relazione ai punti di controllo a terra;
- C3\_2 - per verificare la precisione delle nuvole di punti stereoscopiche rispetto alle nuvole di punti laser, considerando il vicino punto di controllo a terra.

In tale contesto, la misura della qualità dei dati è stata associata a uno o più processi di valutazione in termini di DQM (Data Quality Measure) per ciascun DQE (Data Quality Element) di qualità degli stessi. Per il nostro caso di studio sono descritti in dettaglio due DQM (C2\_4 e C2\_5), come esemplificazioni particolari delle procedure da implementare.

Per quanto riguarda C2\_4, l'identificazione di un metodo oggettivo per evidenziare anomalie nella valutazione della luminanza ha presentato alcuni aspetti critici. In particolare, la percentuale di immagini in cui l'esposizione presentava difficoltà di allineamento così come di colorazione delle nuvole era piuttosto elevata. Quindi, poiché ogni lotto comprende circa 250.000 immagini, è importante stabilire un livello di confidenza di qualità a priori, per evitare l'elaborazione di tutti i dataset. Per quanto riguarda la qualità dell'immagine, la procedura adottata trasforma il colore di ogni pixel in una intensità equivalente di grigio, quindi costruisce l'istogramma dei valori di grigio usando una scala di 256 classi. Come indice caratteristico della distribuzione dei grigi considera infine il valore della mediana, che viene assunto come valore di luminanza per l'immagine. Vengono classificate anomale, ovvero eccessivamente sovra o sottosposte, le immagini, il cui valore di luminanza è esterno a  $k$  volte la deviazione standard rispetto alla mediana delle luminanze complessive del campione. La definizione dell'ampiezza  $k$  e della percentuale  $p$  di immagini anomale che possono causare il rifiuto del dataset è uno dei principali aspetti da modulare. Poiché non espressa in termini di requisiti nelle specifiche del prodotto dati, questa scelta è stata fatta dal verificatore imponendo  $k=2.5$  e  $p=0.03$ . Le considerazioni che inducono a fissare soggettivamente il parametro soglia  $k$  a 2.5 sono costituite dal fatto che in una distribuzione normale standard, il 99% circa dei campioni

ricade appunto nell'intervallo centrato sulla media (mediana) con raggio 2.5 volte la deviazione standard (SD). Il valore massimo di errore ammissibile nel rilievo cartografico è stato inoltre fissato dal committente alla soglia del 3%, per questa ragione si è ritenuto ragionevole estenderne la validità anche all'acquisizione delle immagini, da cui l'impostazione di  $p$  a 0.03.

Un altro aspetto critico era legato alla valutazione dell'intero dataset, tenendo conto dei tempi di elaborazione dovuti a questo controllo. Per ovviare a questo problema, secondo la valutazione SampleBasedInspection (ISO19157, 2013), è stata eseguita una procedura statistica, riducendo il più possibile il campione ma mantenendo comunque una valutazione affidabile. Per il caso in esame si è assunto che la percentuale di errore in un campione ridotto con un'ampiezza  $N$  sia la stima dell'errore reale sconosciuto. Pertanto, l'intervallo confidenziale è stato calcolato con livello di significatività ( $\alpha$ )=0.01 ed una validazione positiva è stata concessa per un valore di stima inferiore a  $p$  e limite superiore nell'intervallo di confidenza inferiore a 0.06, ovvero il doppio dell'errore reale accettabile  $p$ . Inoltre, in questo caso, è sembrato ragionevole stabilire un'ampiezza del campione di  $N = 300$  come un compromesso tra il tempo di elaborazione e l'intervallo di ampiezza accettabile.

Per quanto riguarda C2\_5, il problema principale era legato alla dimensione dei dati da convalidare. Ciascun tronco di nuvole di punti può contenere oltre 20 GB e numero complessivo di tronchi per ogni fornitura generalmente superiore a 200. Oltre a ciò, è stato necessario implementare la validazione automatica della contiguità tra le nuvole di punti. Per il caso studio, sono stati calcolati i tronchi di confine di ogni corsa e sono state considerate le zone di reciproca sovrapposizione (FootPrints), quindi è stato scelto un punto interno per il confronto rispetto alle nuvole di punti in cui erano contenute. Questa valutazione è stata effettuata utilizzando un buffer cilindrico con raggio  $R$  e altezza  $h$  ed applicando l'algoritmo Iterative Closest Point (ICP). La convalida è accettabile se l'RMS è inferiore alla soglia  $e$ . I valori dei parametri selezionati erano  $R=0.5$  m,  $h=2.0$  m ed  $e=0.10$  m. Opzionalmente, è stata effettuata una valutazione tra densità corrispondenti, imponendone una soglia limite.

Come conclusione intermedia per questa fase, la valutazione del rilievo terrestre potrebbe essere espressa in termini di aspetti critici relativi all'identificazione del DQE e alla modellizzazione del DQM correlato.

La scelta di adottare modelli statistici di campionamento ha avuto un impatto decisivo sulla riduzione delle elaborazioni necessarie al completamento delle attività di validazione, soprattutto nei casi in cui la numerosità degli elementi da analizzare assumeva le dimensioni tipiche dei contesti BIG Data.

#### 4.5 La validazione del database

In termini di validazione del database, sono stati considerati sia aspetti specifici che validazioni riferibili all'intero processo, dato che l'acquisizione dei contenuti è ottenuta da diverse fonti, in momenti diversi, raccogliendo dati esistenti e monitorandoli e armonizzando oggetti topografici di base (DBGT) con oggetti tematici specifici dell'ambito ferroviario.

La validazione formale e la certificazione di qualità dei prodotti cartografici è stata eseguita seguendo le specifiche di prodotto dei dati ed in conformità agli standard internazionali (ISO19131, 2018 e ISO19157, 2013).

Le specifiche di prodotto definiscono quelle caratteristiche, attributi e relazioni che sono considerati importanti e devono

essere garantiti congruenti nel dataset di fornitura. Il DBGT è implementato secondo la metodologia del GeoUML (<http://geo.spatialdbgroup.polimi.it/en/>). La metodologia GeoUML è supportata dagli strumenti GeoUML, tra questi sono stati considerati:

- GeoUML Catalogue, che supporta la definizione di schemi concettuali (specifica) e la specifica di alcuni parametri utilizzati nella generazione di schemi fisici in base ai modelli di implementazione scelti.
- GeoUML Validator, che esegue il controllo di conformità di un dataset o di un database rispetto a una specifica (schema concettuale) prodotta dal GeoUML Catalogue.

Le relazioni di base del database e i vincoli topologici all'interno di una classe o tra classi diverse sono stati autocertificati dalla ditta esecutrice tramite l'uso del GeoUML Validator (Fig. 4) e documentati su specifici report.

Elenco dei vincoli GeoUML controllati
SCCONSTRAINTNAME
( Sup_sede.uso = "autostradale"OR Sup_sede.uso = "stradale") PONTE.Supe.superficie ( INT) esiste ( AC_VEI.SottoareeDi_Sede( Sede = "su ponte/viadotto/cavalcavia" ) , AR_VMS.SottoareeDi_Sede( Sede = "su ponte/viadotto/cavalcavia" ) )
( Sup_sede.uso = "ferroviario") PONTE.Supe.superficie ( INT) esiste SD_FER.SottoareeDi_Sede( Sede = "su ponte/viadotto/cavalcavia" )
( Sup_sede.uso = "pedonale") PONTE.Supe.superficie ( INT) esiste AC_PED.SottoareeDi_Sede( Sede = "su ponte/passerella pedonale" )
( uso = "autostradale"OR uso = "stradale") GALLER.Supe.superficie ( INT) esiste ( AC_VEI.SottoareeDi_Sede( Sede = "in galleria" ) , AR_VMS.SottoareeDi_Sede( Sede = "sotterraneo" ) )
( uso = "ferroviario") GALLER.Supe.superficie ( INT) esiste SD_FER.SottoareeDi_Sede( Sede = "in galleria/sottopassaggio pedonale" )
( uso = "pedonale") GALLER.Supe.superficie ( INT) esiste AC_PED.SottoareeDi_Sede( Sede = "in galleria/sottopassaggio pedonale" )
@_AB_CDA.Estensione BND compostoDa AB_CDA.trattiContorno3DDi_Tipo_sponda()
@_AB_CDA.Estensione PLN compostoDa AB_CDA.SottoareeDi_Livello()
@_AC_CIC.Estensione PLN compostoDa AC_CIC.SottoareeDi_Fondo()

Figura 4 - Vincoli topologici e GeoUML Validator

Per quanto riguarda la scelta dei metodi di campionamento per la valutazione, è stato preso in considerazione l'allegato E dello standard sulla qualità dei dati geografici (ISO19157, 2013). In particolare, per quanto riguarda il processo di valutazione della qualità, dopo la definizione dell'unità di qualità dei dati, sono state prese in considerazione le seguenti fasi:

- specificazione delle misure di qualità dei dati: coerenza logica, completezza;
- specificazione delle procedure di valutazione della qualità dei dati;
- determinazione dell'output di valutazione della qualità dei dati: identificazione di errori, coerenza logica, completezza, accuratezza tematica (quantitativa, qualitativa).

Pertanto, in base alle specifiche di prodotto dei dati, oltre ai controlli di autocertificazione di qualità interni, condotti dalla Ditta esecutrice, sono stati eseguiti ulteriori specifici processi di validazione orientati a:

1. Verifica relativa al database nel suo insieme:
  - a. Validazione degli aspetti geometrici in termini di congruenza tra le geometrie e quelli previsti nelle specifiche di prodotto come dimensioni minime di aree e linee in base al livello di dettaglio;
  - b. Validazione della topologia dell'oggetto che appartiene a zone diverse (Fig. 1) con connessione di congruenza 3D lungo i confini;
2. Verifica con adozione di metodi di campionamento:
  - a. Accuratezza tematica e di completezza: basata sul fatto che ogni classificazione può essere soggetta ad errori di omissione/commissione e che la completezza o l'accuratezza tematica hanno sempre un certo grado di soggettività. Ad esempio, la classificazione errata di una casa come edificio industriale potrebbe alternativamente essere considerata come un errore di omissione dell'una e commissione dell'altra;

b. Precisione di misurazione: valutazione del posizionamento dell'oggetto in base a diverse accuratèzze, ovvero 0.40 m per scala 1:1.000 e 0.60 m per scala 1:2.000.

Per le verifiche di cui al punto 1) sono state definite delle procedure ad hoc eseguite da script in linguaggio Python<sup>1</sup>, mentre le fasi dal punto 2) sono state eseguite manualmente, ripetendo l'acquisizione aerofotogrammetrica di un campione estratto casualmente dal dataset.

Il limite di accettazione è stato definito sulla base di criteri di valutazione ormai consolidati nella qualità dei prodotti cartografici e associati ad altri più "sperimentali", con riferimento a quanto definito in For inspection by sampling di cui all'allegato F dello standard sulla qualità dei dati geografici (ISO19157, 2013) e dello standard sui metodi statistici (ISO3534-2, 2006).

Inoltre, con riferimento agli operatori spaziali nei GIS, la mancanza di strumenti topologici 3D, sia nei software open source che in quelli commerciali, ha richiesto una fase di validazione ad hoc della terza coordinata, prestando attenzione alle consistenze tra oggetti diversi tramite l'implementazione specifica di procedure informatiche. Questo aspetto ha avuto un impatto significativo sulla pianificazione delle risorse per la modellazione e l'implementazione dell'infrastruttura geografica di interesse.

Uno dei principali vincoli topologici critici all'interno di un DBGT si riferisce alla valutazione delle congruenze tra diversi modelli topologici come la topologia areale rispetto alla topologia a grafo. Il primo si basa sugli operatori topologici spaziali (Egenhofer et al., 1994 e Clementini et al., 1996) per oggetti topografici di base, mentre il secondo si basa sul sistema di riferimento lineare e sulla topologia di un *network* (INSP.D2.10.1, 2013 e ISO19148, 2012), adatto alla definizione delle reti di trasporto. Nel contesto delle ferrovie, la mezzeria del binario ferroviario deve trovarsi all'interno della massicciata ferroviaria (tenendo conto delle eccezioni, ad esempio passaggi a livello, ecc.). Inoltre, la mezzeria del binario nella terza dimensione deve avere valori congruenti con la terza dimensione del ballast ferroviario (Fig. 5).

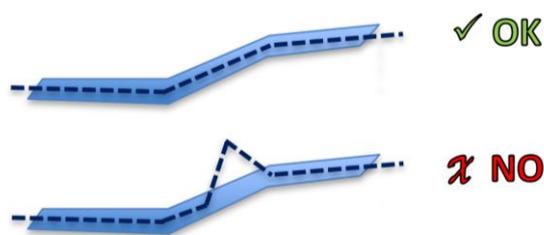


Figura 5 - Vincolo topologico 3D

## 5. DISCUSSIONE E ASPETTI CRITICI

Sulla base dei risultati delle fasi di validazione, si ritiene utile evidenziare, in considerazione di sviluppi futuri di ricerca, alcune questioni aperte, a partire dal caso studio considerato, le quali saranno oggetto di futuri approfondimenti.

### 5.1 Quantità di dati/validazione di BIG Data

Uno dei principali aspetti critici, che ha un impatto significativo sul processo di validazione, è la quantità di dati che devono essere elaborati con la stessa metodologia e con la capacità

<sup>1</sup> <https://www.python.it/>

(hardware e software) di perseguire risultati utili in un tempo sostenibile. Nell'era digitale dei BIG Data, la disponibilità di strumenti all'interno dei processi del prodotto promuove un nuovo ruolo delle attività di validazione.

Partendo dalla consapevolezza che il limite inferiore dei BIG Data si sta muovendo costantemente verso l'alto, emergono alcuni spunti di discussione. In generale, i BIG Data includono informazioni provenienti da fonti diverse, spesso prive di struttura, ma spesso fornite in tempo reale o quasi in tempo reale, consentendo di valutare il comportamento statistico dei fenomeni nel tempo. L'enorme quantità di dati non è facilmente gestibile dai processi tradizionali e richiede algoritmi e metodologie specifici per fornire i risultati in tempi ragionevoli. Per una singola fornitura di dati geografici, l'associazione di BIG Data è in realtà inappropriata se l'elaborazione non è richiesta in tempo reale, tuttavia nel caso di un rilievo multisensore e per un territorio esteso, i problemi possono essere considerati quasi simili.

Nel caso di un rilevamento di una infrastruttura di trasporto, l'estensione potrebbe essere considerevole e il piano di acquisizione deve essere puntuale, rigoroso e il più breve possibile, al fine di identificare una data per ciascuna tipologia di rilevamento topografico e conseguentemente pianificare la tempistica di aggiornamento. La produzione di dati ha tempistiche piuttosto ristrette, quindi, anche se ogni lotto non può essere classificato come BIG Data, lo diventa in considerazione della frequenza di fornitura. Peraltro, anche nel caso di un livello di dettaglio molto elevato, anche un singolo lotto potrebbe rientrare nel concetto di BIG Data.

Dal punto di vista della certificazione di qualità, queste premesse implicano l'adozione di una strategia che bilanci la richiesta di analizzare grandi quantità di dati con la necessità di produrre risultati in breve tempo. Pertanto, sono stati presi in considerazione metodi statistici per identificare la coerenza del campione più piccolo in grado di fornire una valutazione affidabile dell'intero dataset (con limite di confidenza superiore al 95%).

## 5.2 Definizione semantica/tematica diversa della validazione degli oggetti

I modelli di rete ferroviaria esistenti rappresentano servizi con un elevato valore semantico, ma i loro componenti non interagiscono o hanno relazioni esplicite con le caratteristiche informative di un prodotto cartografico di base. Poiché le reti ferroviarie sono descritte con una semantica ed un dettaglio piuttosto elevato, ben si prestano all'uso di diverse applicazioni: alcune sono dedicate all'uso quotidiano nelle società di servizi pubblici, alcune vengono utilizzate come modelli di scambio di dati e altre rappresentano reti di servizi all'interno di un ambito urbano. Ogni rete rappresenta un'astrazione e una riduzione di un sistema (modello) in sé; tuttavia, non forniscono connessioni immediate al contesto territoriale più ampio, pertanto non sono direttamente spendibili in termini di analisi o simulazione urbana, gestione dei rischi e dei disastri e gestione del ciclo di vita della città. Un modello adatto a tali scopi dovrebbe soddisfare i seguenti requisiti e dovrebbe rappresentare una generalizzazione ammissibile e sottoinsieme di situazioni tali per cui:

- Gli elementi di tale modello devono avere relazioni sia funzionali che strutturali l'uno con l'altro;
- Il modello deve rappresentare elementi indipendenti ma correlati per consentire la simulazione e l'analisi complessa;
- Il modello deve essere valido per diversi tipi di sistemi di rete eterogenei;

- Il modello deve ridurre la complessità preservando le informazioni richieste per l'uso in simulazioni, analisi, calcoli e visualizzazione cartografica in casi di catastrofe.

È necessario definire l'ambiente di modellazione rendendo esplicite le funzionalità pertinenti e le loro relazioni reciproche, consentendo così la modellazione topografica 3D di intere reti, sottoreti e funzionalità di rete, nonché le loro rappresentazioni grafiche. Il conseguente trattamento delle funzionalità di rete come un'astrazione di oggetti del mondo reale (punto di vista topografico) e di un oggetto grafico rappresentato dal proprio grafo dell'infrastruttura, rende il modello flessibile, come accaduto per i modelli creati nei sistemi di gestione GIS per le utilities (Adolphi et al. 2013).

L'aspetto innovativo di questa ricerca riguarda la volontà di andare oltre i vari programmi di conversione dati, verso un ambiente in cui i dati siano veramente integrati, il framework di modellazione sia più universale e gli standard dei dati si estendono su programmi software e piattaforme specifiche adottate da fornitore di dati (Zlatanova et al., 2005).

## 6. CONCLUSIONI

Un'analisi approfondita è stata effettuata sia sui dati forniti in forma di prototipo sia sui primi lotti consegnati ufficialmente, al fine di porre in evidenza gli aspetti metodologici e le questioni critiche descritte in questo articolo.

Gli aspetti salienti che sono stati impattanti rispetto alla validazione sono quelli di seguito elencati:

- La notevole varietà dei dati forniti (immagini, rilievi topografici, tracce GPS, triangolazione aerea, modelli numerici, acquisizione vettoriale, DBGT, immagini panoramiche e telemetria laser ad alta risoluzione), ciascuno con le proprie peculiarità, richiedono capacità e competenze di alto livello per poter essere analizzati e validati;
- L'armonizzazione tra diversi processi di validazione ha riguardato sia fasi ben note come best practices nella rilevazione fotogrammetrica tradizionale, come anche metodi di campionamento applicati per la validazione di dataset "di nuova generazione" (es. da MMS) ma anche per la validazione di DBGT complessi;
- L'enorme quantità di dati appartenenti a ciascun lotto di fornitura;
- La frequenza di rilascio dei lotti e la conseguente necessità di produrre gli output di validazione in poco tempo ma con un alto livello di affidabilità;
- L'impossibilità di eseguire sequenzialmente le operazioni di controllo per motivi formali;
- L'impossibilità di eseguire ricognizioni sul campo a causa dell'estensione spaziale dell'intero progetto di rilevamento.

Per far fronte a tutte le fasi di validazione, innanzitutto è stato istituito un gruppo di lavoro, riunendo competenze specifiche per ciascuna fase d'acquisizione/elaborazione dei dati. Il secondo passo è stato quello di identificare i punti di connessione tra le fasi per condividere gli output risultanti, permettendo così a ciascun gruppo di lavoro di concentrarsi solo sui propri problemi specifici, ma avendo allo stesso tempo la piena fiducia nell'affidabilità dei test di controllo effettuati sugli altri processi del sistema di validazione.

Inoltre, sono state sviluppate procedure specifiche per automatizzare il processo per tutte quelle attività in cui la validazione potrebbe essere condotta da una serie di operazioni in cascata e il relativo esito potrebbe essere stabilito sulla base di un confronto con soglie prestabilite. In quei casi, quando la fase di certificazione prevede operazioni manuali o la quantità

di dati è così grande che anche le procedure automatiche sono troppo onerose in termini di tempo macchina, sono state identificate metodologie adeguate di selezione del campione minimo significativo su cui effettuare un test affidabile.

I metodi di campionamento sono stati quindi utilizzati per sviluppare le procedure e bilanciare i tempi necessari per il completamento delle operazioni di controllo.

Ulteriori perfezionamenti sono stati infine apportati alla fornitura dei primi lotti, dopo l'implementazione delle metodologie sopra citate, per rendere più semplici e solide le procedure anche in presenza di dati non completamente conformi.

Il processo di validazione applicato sul caso studio, anche se relativo alle infrastrutture ferroviarie, si ritiene possa essere applicato in modo congruente a ciascun database spaziale in cui le fonti di rilevamento siano multiple purché in termini di accuratezza spaziale e temporale. Oggigiorno, accade sempre di più grazie all'evoluzione tecnologica, che per gestire complesse infrastrutture di dati spaziali, il requisito principale nella rilevazione dei dati, sia che questi siano continuamente aggiornati/aggiornabili, con fonti diverse nel miglior modo possibile. Pertanto, la correttezza dell'approccio, indipendentemente dallo specifico livello tematico spaziale d'interesse, è stata supportata dall'adozione agli standard internazionali relativi alla specifica di prodotto (ISO19131, 2018) ed alla qualità (ISO19157, 2013) dei dataset geografici validati.

## 7. RINGRAZIAMENTI

Per la collaborazione fornita gli autori desiderano esprimere la loro gratitudine al Generale Enzo Santoro per il contributo fornito al presente articolo. Gli autori desiderano ringraziare anche la rete ferroviaria italiana (RFI) per aver reso disponibili i dataset MUIF per questo caso di studio.

## 8. RIFERIMENTI

Adolphi, T., Nagel, C., Heinrich, T., Kolbe, T., 2013, Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization. DOI: 10.1007/978-3-642-29793-9\_3 - In book: Progress and New Trends in 3D Geoinformation Sciences Project: BIM and GIS integration.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A., 2015, Applications of 3D City Models: State of the Art Review. ISPRS International Journal of Geo-Information, 4(4): 2842-2889.

Carrion, D., Maffei, A., Pinto, L., 2008, Metodi e modelli per il controllo di qualità di data base topografici multiscala. Atti 12a Conferenza Nazionale ASITA, 21 -24 ottobre 2008, L'Aquila, IT, ISBN 978-88-903132-1-9.

Clementini, E., Di Felice, P., 1996, A model for representing topological relationships between complex geometric features in spatial databases. Inform. Syst. 90, 1-4, 121-136.

Corongiu, M., Tucci, G., Santoro, E., Kourounioti, O., 2018, Data Integration Of Different Domains In Geo-Information Management: A Railway Infrastructure Case Study. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLII-4, 121-127, <https://doi.org/10.5194/isprs-archives-XLII-4-121-2018>.

DM, 2012, Decreto Ministeriale del 10 November 2011 (2012), Regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici. (12A01800) (GU Serie Generale n.48 del 27-02-2012 - Suppl. Ordinario n. 37).

Egenhofer, M. J., Clementini, E., Di Felice, P., 1994, Topological relations between regions with Holes. Int. J. Geograph. Inform. Syst. 8, 2, 128-142.

Kolbe, T., 2009, Representing and Exchanging 3D City Models with CityGML. In: 3D Geo-Information Sciences pp 15-31 Lecture Notes in Geoinformation and Cartography, Springer Berlin Heidelberg.

INSP., 2007, Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>.

INSP.D2.8.I.7, 2014, INSPIRE D2.8.I.7 Data Specification on Transport Networks, Technical Guidelines INSPIRE Infrastructure for Spatial Information in Europe <https://inspire.ec.europa.eu/id/document/tg/tn>.

INSP.D2.10.1, 2013, Data Specifications, Base Models, Generic Network Model <https://inspire.ec.europa.eu/documents/inspire-data-specifications-%E2%80%93-base-models-%E2%80%93-generic-network-model>.

ISO19131, 2018, ISO/CD 19131 Geographic information, Data product specifications. [www.isolutions.iso.org](http://www.isolutions.iso.org).

ISO19148, 2012, ISO/IS 19148, Geographic information, Linear Referencing. [www.isolutions.iso.org](http://www.isolutions.iso.org).

ISO19157, 2013, ISO/FDIS 19157 Geographic information, Data quality. [www.isolutions.iso.org](http://www.isolutions.iso.org).

ISO3534-2, 2006, ISO 3534-2 Applications of statistical methods Statistics, Vocabulary and symbols — Part 2: Applied statistics. [www.isolutions.iso.org](http://www.isolutions.iso.org).

Thaduri, A., Galar, D., Kumar, U., 2015, Railway assets: A potential domain for big data analytics. Procedia Computer Science, 2015 INNS Conference on Big Data, Volume 53, 2015, Pages 457-467 doi: 10.1016/j.procs.2015.07.323.

Zlatanova, S., Prosperi, D., 2005, Large-scale 3D Data Integration. CRC Press Reference. ISBN 9780849398988 - CAT# 9898.