

**Gestione automatizzata delle Specifiche
di una Base Dati Territoriale orientata alla gestione e condivisione dei dati
Inquadramento Generale**

22 settembre 2008

Redazione: Alberto Belussi, Federica Liguori, Jody Marca, Sara Migliorini, Mauro Negri, Giuseppe Pelagatti (coordinatore)

mail: SpatialDBgroup@elet.polimi.it

Allegato n. 1 alla Relazione Finale del progetto di Ricerca "Database Topografici" finanziato dal Centro Interregionale

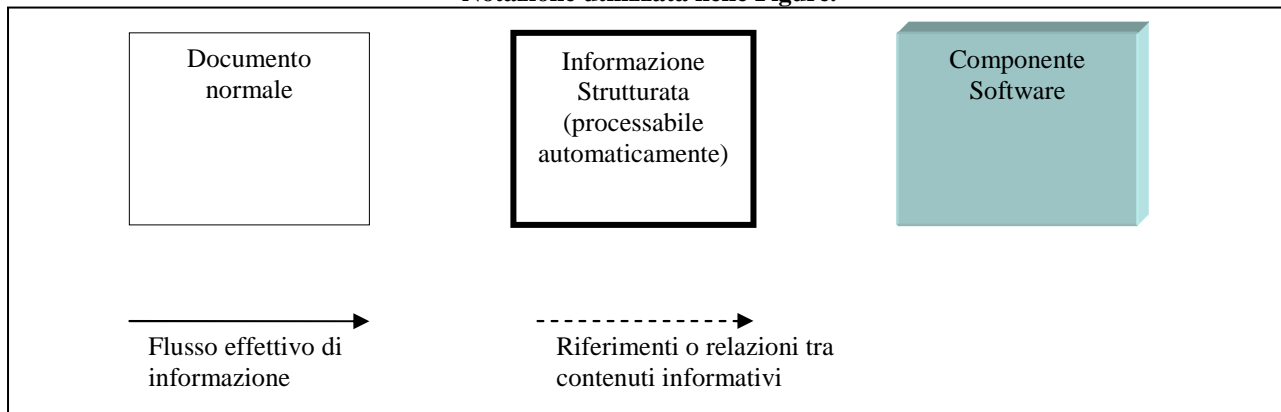
CONTENUTO

1. INTRODUZIONE
2. SPECIFICHE DI CONTENUTO E GEOUML CATALOGUE
 - 2.1. Specifiche di contenuto
 - 2.2. GeoUML catalogue
 - 2.3. Le funzioni di gestione di una specifica di contenuto
3. CONFORMITA' DI UN DATASET A UNA SPECIFICA DI CONTENUTO
 - 3.1. Conformità Reale e Conformità Intrinseca
 - 3.2. GeoUML validator
4. DEFINIZIONE DI SPECIFICHE DI CONTENUTO STANDARD
 - 4.1. Obiettivi
 - 4.2. Compatibilità
 - 4.3. Coerenza semantica
 - 4.4. Uso combinato di Coerenza e Compatibilità
 - 4.5. Livelli ulteriori di flessibilità
5. I MODELLI IMPLEMENTATIVI E LA PRODUZIONU DI SCHEMI FISICI
 - 5.1. Funzione del Modello Implementativo
 - 5.2. Generazione di un Application Schema di alto livello
 - 5.3. Generazione dello Schema GML_ESF di una SC
 - 5.4. Altre categorie di Modelli Implementativi
6. STRUMENTI SOFTWARE E MODELLO IMPLEMENTATIVO

Appendice A1. RAPPORTO TRA GEOUML METHODOLOGY E GLI STANDARD ISO TC211

Appendice A2. SPECIFICHE INTESAGIS E GEOUML METHODOLOGY

Notazione utilizzata nelle Figure.



1. INTRODUZIONE

Questo documento propone delle linee guida metodologiche basate su una adeguata strumentazione software per portare avanti il processo di realizzazione dei DataBase Topografici secondo la linea definita dalle Specifiche IntesaGIS nel 2003 (rivista nel 2006).

Le specifiche IntesaGIS hanno infatti lasciato aperti alcuni aspetti fondamentali:

1. Come derivare dalle Specifiche di Contenuto (a livello concettuale) gli schemi fisici adeguati alle diverse esigenze (gestione del Database Topografico, trasferimento di dataset, ecc...)
2. Come verificare la conformità di un Dataset a una Specifica data
3. Come garantire le premesse per un'efficace condivisione dei dati pur permettendo ai singoli enti di ritagliare le specifiche secondo le loro esigenze
4. Come garantire le premesse per un'efficace condivisione dei dati e del software pur permettendo ai singoli enti di adottare diverse tecnologie/formati fisici per memorizzare i loro Dataset

Durante i 5 anni di sperimentazione trascorsi dal 2003 i diversi enti hanno risposto in maniera variegata a queste esigenze; tra l'altro, hanno definito capitolati che ritagliavano le specifiche e definivano strutture fisiche in maniera diversa uno dall'altro e hanno realizzato procedure di controllo per verificare la conformità dei Dataset prodotti alle specifiche. Il risultato di questo modo di procedere in ordine sparso è costituito da un'oggettiva difficoltà a stabilire quali siano i contenuti realmente condivisi e quali i formati di scambio adottabili; inoltre, non si garantisce il riuso degli investimenti fatti in software di controllo.

Questo documento propone un Metodo di Gestione delle Specifiche e di Condivisione dei Dati, supportato dallo sviluppo di alcuni strumenti Software, orientato a garantire agli enti la massima possibile autonomia nella costruzione delle loro Basi Dati Territoriali, ma, allo stesso tempo, a garantire il massimo grado di condivisione dei dati e di riuso del software.

Lo strumento software fondamentale nell'impostazione proposta è il **GeoUML catalogue**, che permette di gestire una Specifica di Contenuto e, tramite opportune estensioni integrabili nello strumento base (plug-in), di associare a tale specifica uno o più **Schemi Fisici**. Le regole in base alle quali uno schema fisico viene prodotto da una Specifica di Contenuto sono dette **Modello di Implementazione**.

Allo stato attuale (versione 2008), GeoUML catalogue è in grado di produrre solamente il modello di implementazione per una specifica struttura fisica basata sul linguaggio GML, detta GML_ESF.

Lo strumento GeoUML catalogue (versione 2008) consente di affrontare alcuni dei problemi citati sopra, in particolare:

1. Gestire l'evoluzione di una Specifica, mantenendo consistenti tutti i documenti che la definiscono a diversi livelli
2. Definire diverse Specifiche in modo che siano garantite alcune importanti relazioni di corrispondenza tra di loro
3. Derivare da una Specifica il corrispondente schema fisico in formato GML_ESF

Le funzioni svolte da GeoUML catalogue possono essere considerate relativamente complete per quanto riguarda la gestione delle Specifiche di Contenuto in quanto tali; per quanto riguarda la generazione degli schemi fisici GeoUML catalogue (versione 2008) fornisce un ambiente estendibile tramite il quale è possibile produrre estensioni (plug-in) in grado di generare diversi tipi di schemi fisici a costi assai contenuti, coprendo in tal modo tutte le esigenze di gestione delle specifiche dal livello concettuale a quello fisico. In particolare, GeoUML catalogue (versione 2008) è dotato già di un plug-in per produrre un XSD per il formato GML_ESF, che viene proposto come formato di scambio tra gli enti.

Oltre ad avere una sua utilità intrinseca, la gestione delle specifiche tramite GeoUML catalogue costituisce, da un punto di vista più generale, anche il punto di partenza di una metodologia per affrontare in maniera sistematica i problemi del processo complessivo di gestione dei Database Topografici e della loro condivisione. Chiameremo **GeoUML methodology** tale metodologia, che sostanzialmente mira a trasferire nel contesto dell'informazione territoriale molti principi consolidati delle metodologie applicate da lungo tempo nel campo delle Basi Dati tradizionali, il più fondamentale dei quali è sicuramente *il principio della separazione tra gli aspetti di contenuto e gli aspetti legati alla tecnologia*; tale principio è ovviamente applicato in GeoUML catalogue supportando la separazione tra la specifica di contenuto e i relativi schemi fisici. L'applicazione di questo principio richiede però un processo di formazione degli operatori del settore, perché la tendenza a definire un contenuto informativo direttamente tramite una implementazione fisica è molto diffusa e apparentemente semplificatrice.

Diversi studi e sperimentazioni hanno permesso di verificare che la chiara separazione tra i due aspetti permette di migliorare alcuni processi fondamentali, come la produzione di specifiche per la produzione dei dati, il controllo di qualità di un dataset rispetto a tali specifiche, l'integrazione e armonizzazione di basi dati diverse.

La disponibilità di specifiche in formato processabile permette inoltre lo sviluppo di *strumenti guidati dalle specifiche (Schema Driven)* che operano sui dati, perchè fornisce a tali strumenti una interpretazione del significato dei dati stessi. Ad esempio, si pensi ad uno strumento capace di controllare che un Dataset sia conforme a una specifica data .

Lo sviluppo e l'uso di strumenti di questo tipo non solo non è in conflitto con l'autonomia di scelta dei singoli enti relativamente agli strumenti (commerciali o liberi) utilizzati per creare e gestire i loro DB topografici, ma, al contrario, sostiene tale autonomia perchè supporta gli aspetti di condivisione tra enti senza richiedere che tutti gli enti adottino la stessa tecnologia o piattaforma tecnologica.

Questa finalità è tipicamente quella degli standard; in effetti la GeoUML methodology può essere considerata come una estensione (perfettamente compatibile) degli standard ISO TC211 e OGC, orientata a utilizzare tali standard nel modo più conveniente per supportare la linea definita dalle specifiche IntesaGIS del 2003. Le motivazioni per cui gli standard esistenti da soli non sono adeguati a questo scopo sono spiegate in Appendice Ax.

E' opportuno richiamare il fatto che la GeoUML methodology affronta esclusivamente il problema di metodo e di strumenti per la gestione e condivisione dei Database Topografici ed è indipendente dagli aspetti di contenuto esempi di problemi di metodo sono:

- come scrivere le Specifiche che descrivono il Contenuto di un Dataset
- come verificare che un Dataset sia effettivamente conforme ad una data Specifica
- come garantire che una Specifica adottata da un Ente sia adeguata a supportare processi di condivisione dei dati a livello superiore
- ecc...

esempi di problemi di contenuto sono:

- quali classi di oggetti devono essere presenti in un Dataset
- quali attributi devono possedere gli oggetti di una classe
- quale tipo di componenti spaziali devono possedere gli oggetti di una classe
- quali vincoli di integrità devono soddisfare le componenti spaziali

ecc...

Da un punto di vista normativo gli enti che esercitano un'autorità di regolamentazione possono definire norme sia rispetto al metodo che al contenuto; tuttavia è opportuno che i due aspetti siano mantenuti separati; in generale si può dire che:

- la definizione di uno stesso metodo può essere applicata a diverse Specifiche di Contenuto
- è difficile dare una valenza precisa a una Specifica di Contenuto in assenza di un minimo di regole relative al metodo col quale le Specifiche devono essere definite ed applicate

Relazione tra la GeoUML methodology e le specifiche IntesaGIS

Le Specifiche di Contenuto IntesaGIS prodotte nel 2004 e 2006 (documenti 1007_1-2 e 1007_4) vengono acquisite e inserite nel GeoUML catalogue (e costituiscono una Specifica di Contenuto rispetto alla quale è possibile verificare automaticamente opportune relazioni di conformità da parte di altre specifiche, come descritto nel capitolo 4 di questo documento).

2. SPECIFICHE DI CONTENUTO E GEOUML CATALOGUE

2.1 Specifiche di Contenuto

Una Specifica di Contenuto (SC) descrive il contenuto informativo di un Database per mezzo di diversi tipi di definizioni che hanno scopi diversi:

1. definizione degli **elementi informativi**, cioè degli elementi che devono essere rappresentati nel Database, (classi, attributi, attributi geometrici, associazioni, domini) in maniera indipendente dal modo in cui tali elementi sono rappresentati in una particolare tecnologia di memorizzazione dei dati
2. definizione dei **vincoli di integrità**, cioè delle proprietà intrinseche che gli elementi informativi devono soddisfare (con proprietà intrinseche si intendono le proprietà verificabili sugli elementi informativi stessi, senza osservazioni dirette del mondo reale)
3. **elementi descrittivi**: in questa categoria rientrano tutte le informazioni utilizzabili dagli esseri umani per capire come interpretare un contenuto del database in termini di realtà rappresentata e viceversa come interpretare una situazione reale nei termini degli elementi informativi (ad esempio, la modalità di rilievo)

L'insieme costituito dagli elementi informativi e dai vincoli di integrità costituisce la **Parte Strutturata** della Specifica di Contenuto, mentre gli elementi descrittivi costituiscono la **Parte Informale** della specifica.

Le regole utilizzate per definire la parte strutturata delle specifiche costituiscono il **modello GeoUML**; la definizione precisa di tali regole è fornita nel documento "*Il modello GeoUML (versione 2008)*"

2.2 GeoUML catalogue

Per creare, modificare e visualizzare una Specifica di Contenuto si utilizza il **GeoUML catalogue**; tale applicazione software verrà resa disponibile secondo modalità in fase di definizione.

L'applicazione è interamente scritta in linguaggio Java e può essere quindi utilizzata su qualsiasi piattaforma.

Il GeoUML catalogue può esportare/importare le specifiche in un formato XML pubblico precisamente definito tramite un XSD, quindi è possibile realizzare strumenti capaci di operare sulle specifiche gestite da GeoUML catalogue (inclusi altri gestori delle specifiche sostitutivi di GeoUML catalogue). Nel seguito indicheremo una specifica SC in tale formato con la notazione SC.xml.

L'XSD che descrive tale formato verrà reso disponibile insieme al GeoUML catalogue.

GeoUML catalogue possiede due tipi di funzionalità

1. Funzioni per la gestione delle Specifiche di Contenuto: le funzioni di questo tipo sono descritte ad alto livello nel seguito di questo capitolo, e presuppongono una comprensione del modello GeoUML
2. Funzioni per la produzione di Schemi Fisici in base ai corrispondenti Modelli Implementativi: queste funzioni sono trattate nel capitolo 5 e non vengono ulteriormente prese in considerazione per il momento

2.3 Le funzioni di gestione di una Specifica di Contenuto

Le funzioni per la gestione delle Specifiche di Contenuto possono essere a loro volta suddivise nelle seguenti categorie, alle quali sono associate in maniera indicativa varie categorie di utenti (vdi figura 2.1):

1. Funzioni relative alla Identificazione, gestione complessiva, importazione ed esportazione di una specifica nel formato standard (per Amministratore della Specifica)
2. Funzioni di Editing dei contenuti (per utenti abilitati dall'amministratore)
3. Funzioni di Visualizzazione e Ricerca dei contenuti (per utenti generici)
4. Funzioni di preparazione e produzione della documentazione standard (per utenti abilitati dall'amministratore)
5. Funzioni di gestione delle relazioni tra specifiche diverse (per Amministratore della Specifica)

Nel seguito vengono descritte le prime 4 funzioni; per la funzione di gestione delle relazioni tra specifiche diverse si rimanda al capitolo 4.

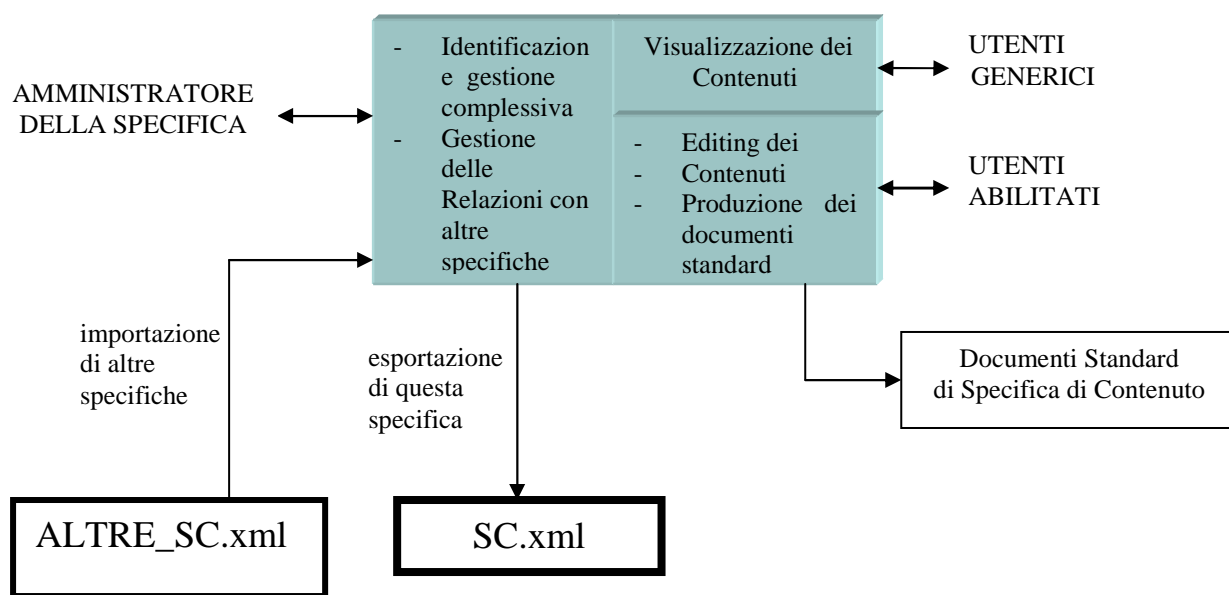


Figura 2.1

3 CONFORMITA' DI UN DATASET A UNA SPECIFICA DI CONTENUTO

3.1 Conformità Reale e Conformità Intrinseca

La conformità di un Dataset D a una specifica di contenuto SC è composta da due aspetti:

1. **Conformità Reale:** la conformità reale riguarda la corrispondenza tra il contenuto informativo del Dataset e la porzione di Mondo Reale alla quale il Dataset si riferisce; gli elementi informativi e i relativi elementi descrittivi della SC determinano come tale corrispondenza debba essere valutata
2. **Conformità Intrinseca:** la conformità intrinseca riguarda la consistenza dell'informazione contenuta nel Dataset; sia gli elementi informativi che i vincoli della SC sono utilizzati per valutare la conformità intrinseca

La valutazione della qualità di un Dataset deve tener conto sia della Conformità Reale che della Conformità Intrinseca alla Specifica di Contenuto.

Nella GeoUML methodology viene presa principalmente in considerazione la conformità intrinseca, perchè solamente quest'ultima può essere valutata automaticamente.

Un Dataset D è intrinsecamente conforme a una Specifica di Contenuto SC se e solo se:

- a. tutti i dati contenuti in D corrispondono a elementi informativi di SC
- b. i dati contenuti in D soddisfano tutti i vincoli di integrità definiti in SC

Si osservi che la condizione (a) non implica che per ogni elemento informativo di SC esistano dei dati corrispondenti; ad esempio, in SC può essere definita una classe della quale non esistono istanze nel Dataset senza violare la Conformità Intrinseca

3.2 GeoUML validator

Chiameremo **GeoUML validator** uno strumento in grado di operare il controllo di conformità intrinseca di un Dataset relativamente ad una qualsiasi Specifica gestita da GeoUML catalogue (vedi figura 3.1).

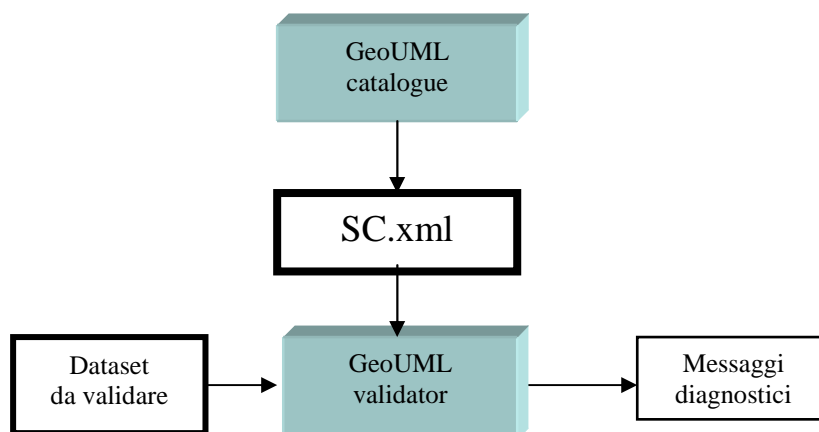


Figura 3.1

Vale la pena di sottolineare che anche le tradizionali operazioni di collaudo, orientate alla verifica della Conformità Reale, possono beneficiare sensibilmente del supporto di un Validatore automatico di Conformità Intrinseca, perchè alcuni errori dei dati rispetto alla Conformità Intrinseca costituiscono spesso dei segnali di possibili errori di Conformità Reale. Ad esempio, un validatore di conformità intrinseca potrebbe segnalare l'assenza di istanze di una classe come avvertimento (Warning) per il collaudatore; la verifica che l'assenza di istanze corrisponda alla situazione del mondo reale rappresentato nel Dataset e non sia invece un errore riguarda la Conformità Reale. L'efficacia di questo tipo di segnalazioni nel migliorare la qualità dei collaudi è stata verificata in numerose occasioni.

4. DEFINIZIONE DI SPECIFICHE DI CONTENUTO STANDARD

4.1 Obiettivi

Questo capitolo della GeoUML methodology tratta il metodo utilizzabile per ottenere una regolamentazione delle Specifiche di Contenuto che vengono utilizzate da diversi Enti per creare le proprie Basi Dati Territoriali.

La valenza istituzionale e giuridica del processo di regolamentazione non interessa la GeoUML methodology; ciò che essa definisce è il meccanismo fondamentale tramite il quale un Ente può esercitare una funzione di regolamentazione.

I concetti alla base di tale meccanismo sono i seguenti:

- definizione di particolari Specifiche di Contenuto dette **Specifiche di Contenuto Standard (SC_standard)**; ad esempio, “National Core” oppure “IntesaGIS 2006”
- definizione di particolari **relazioni** relative alle Specifiche di Contenuto Standard, che le **SC_regolate**, cioè le SC adottate dagli enti soggetti alla regolamentazione, devono soddisfare

Ovviamente la relazione più stringente è l'**identità**; richiedere l'identità con una SC_standard impone di utilizzare una SC_regolata identica alla SC_standard.

Dato che è ampiamente riconosciuta l'esigenza di applicare relazioni più flessibili, che permettano alle SC_regolate di differenziarsi dalle SC_standard, nel seguito sono definite 2 relazioni più deboli dell'identità, dette **compatibilità e coerenza (semantica)**.

Le relazioni di compatibilità e coerenza semantica sono supportate da GeoUML catalogue (versione 2008) sia controllando il processo di editing dei contenuti di una specifica regolata in modo che la relazione sia sempre soddisfatta durante l'editing, sia come verifica complessiva applicabile a due specifiche. Come mostrato nella seguente figura 4.1, il GeoUML catalogue importa a questo scopo le SC_standard di riferimento.

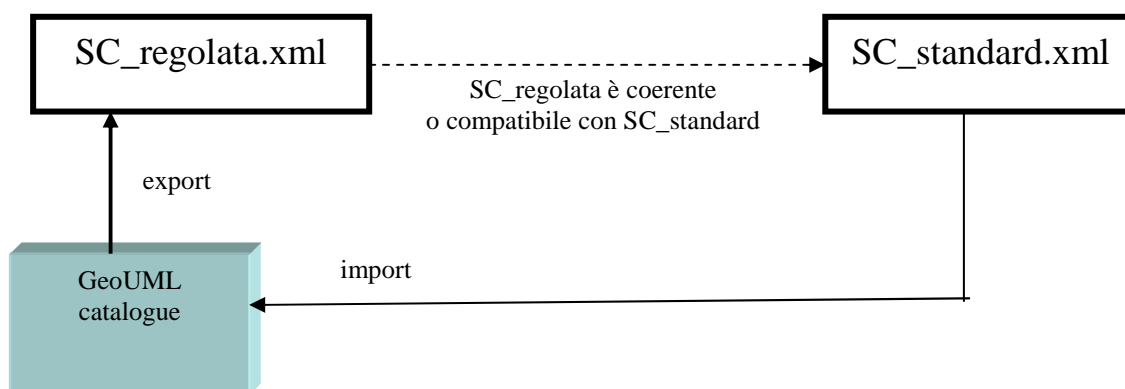


Figura 4.1

4.2 Compatibilità

La compatibilità è definita nel modo seguente: una SC_regolata è compatibile con una SC_standard se la SC_regolata contiene tutti gli elementi informativi, i vincoli di integrità e gli elementi descrittivi presenti in SC_standard.

La relazione di compatibilità permette di garantire l'esistenza di **contenuti condivisi** dalle SC_regolate di tutti gli enti soggetti a tale regola; il contenuto condiviso è ovviamente quello definito dalla SC_standard.

La proprietà di compatibilità garantisce la possibilità di estrarre da un dataset D, conforme a una SC_regolata, a sua volta compatibile con una specifica SC_standard, un Dataset SelD, ottenuto selezionando solo i dati che rappresentano Elementi Informativi di SC_standard. Questo Dataset SelD è conforme (sia in senso Reale che Intrinseco) a SC_standard.

Inoltre, come suggerito dalla seguente figura 4.2, è concettualmente possibile realizzare uno strumento (indicato come “estrattore di dataset compatibili” in figura) capace di estrarre SelD da D automaticamente in base alle due specifiche SC_regolata e SC_standard, che sono state a loro volta prodotte da GeoUML catalogue; inoltre la figura mostra che il GeoUML catalogue utilizzato per produrre la SC_regolata ha importato la SC_standard in modo da poter controllare che la relazione di compatibilità fosse soddisfatta.

Per chiarezza conviene richiamare la sequenza temporale degli eventi sottintesa dalla figura 4.2:

1. Un Ente Regolatore ha definito tramite GeoUML catalogue la specifica SC standard, l'ha congelata ed ha esportato il file SC_standard.xml

2. Un Ente E soggetto a tale regolazione ha creato tramite GeoUML catalogue la specifica SC_regolata, garantendo, sempre tramite GeoUML catalogue, la compatibilità con SC_standard, infine l'ha congelata ed ha esportato il file SC_regolata.xml
3. Un produttore ha prodotto per conto dell'ente E un dataset D, conforme a SC_regolata (tale conformità potrebbe essere verificata tramite GeoUML validator - non riportato in figura)
4. L'ente E è in grado di produrre il Dataset SelD estraendolo dal dataset D tramite l'estrattore di dataset compatibili (l'ente regolatore potrebbe verificare la conformità di SelD con la specifica SC_standard utilizzando GeoUML validator - anche questo non è indicato in figura)

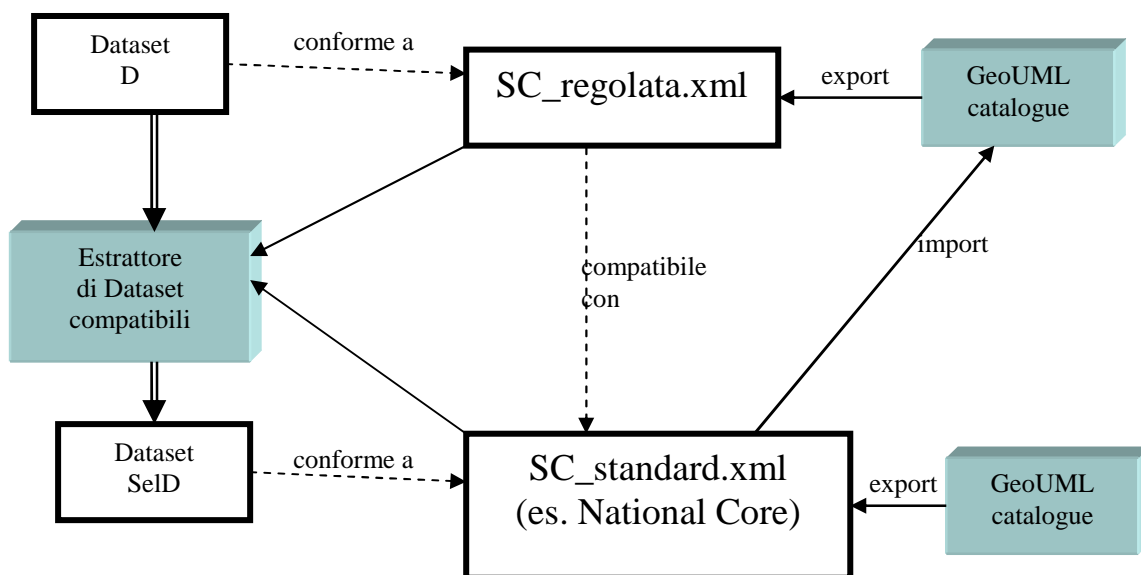


Figura 4.2

4.3 Coerenza Semantica

La coerenza (semantica) mira a ottenere un elevato grado di **omogeneità** tra specifiche di enti diversi che trattano gli stessi dati: significa che nella misura in cui due specifiche riguardano gli stessi contenuti informativi, esse devono essere formulate e strutturate nello stesso modo; tra i vantaggi ottenibili grazie alla omogeneità si possono citare ad esempio:

- risparmi nella creazione delle specifiche stesse e di tutti i derivati (capitolati, ecc...)
- creazione di un linguaggio condiviso tra tutti gli enti nella definizione dei dati, comprensione più accurata e precisa delle specifiche altrui, ecc...
- possibilità di creazione di applicazioni basate sulle specifiche condivise da diversi enti

La coerenza semantica è definita nel modo seguente: una SC_regolata è semanticamente coerente a una SC_standard se valgono le seguenti condizioni:

1. gli elementi informativi di SC_regolata costituiscono un sottoinsieme degli elementi informativi di SC_standard
2. SC_regolata include tutti i vincoli V presenti in SC_standard tali che tutti gli elementi informativi utilizzati nella formulazione di V in SC_standard sono presenti anche in SC_regolata
3. SC_regolata contiene tutti gli elementi descrittivi presenti in SC_standard e relativi agli elementi informativi presenti anche in SC_regolata, eventualmente modificati in modo da tener conto della riduzione dei contenuti

E' importante osservare che le regole elencate non vietano alla SC_regolata di contenere vincoli di integrità aggiuntivi rispetto a quelli presenti nella SC_standard.

La formulazione di diverse SC_regolate coerenti con una stessa SC_standard permette di ottenere un certo grado di omogeneità tra di loro e i relativi vantaggi citati in precedenza. Tuttavia, l'applicazione della Coerenza Semantica a fini normativi è problematica, perché spesso un Ente vuole giustamente aggiungere alla propria SC_regolata degli Elementi Informativi che non sono presenti in SC_standard per gestire aspetti che la SC_standard non prende in considerazione, quindi in questi casi imporre la coerenza come norma vincolante non sarebbe corretto.

Il concetto di coerenza semantica costituisce invece una **raccomandazione** interpretabile nella SC_regolata nel modo seguente: *per tutte le tipologie di Elementi Informativi della SC_regolata presenti nella specifica SC_standard si raccomanda di prelevare esattamente le definizioni presenti in SC_standard, pur non escludendo la possibilità di aggiungere Elementi Informativi che rappresentano tipi di dati non trattati nella SC_standard.*

Possiamo dire che una SC_regolata **massimizza** il proprio livello di coerenza semantica con una SC_standard se gli Elementi Informativi di SC_regolata non derivati da SC_standard si riferiscono esclusivamente a tipologie di dati effettivamente non rappresentabili con gli Elementi Informativi di SC_standard.

La massimizzazione della coerenza semantica non è verificabile automaticamente, perché non è possibile verificare se un Elemento Informativo definito in SC_regolata e non presente in SC_standard corrisponde veramente a un concetto informativo nuovo invece che a una diversa definizione di un concetto già presente in SC_standard; tuttavia è possibile supportare tramite il GeoUML catalogue gli enti che intendono massimizzare la coerenza semantica, come indicato al successivo punto 4.5.

4.4 Uso combinato di Coerenza e Compatibilità

E' possibile regolamentare la creazione di SC_regolate richiedendo contemporaneamente sia la massima coerenza con una SC_standard_1, sia la compatibilità con una SC_standard_2.

In questo modo le SC_regolate devono possedere certi contenuti (compatibilità) e per la parte eccedente tali contenuti devono essere il più possibile omogenee con le specifiche di riferimento (coerenza semantica).

Per rendere possibile la creazione di SC_regolate che siano allo stesso tempo coerenti con una SC_standard_1 e compatibili con una SC_standard_2 è necessario che SC_standard_2 sia a sua volta coerente con SC_standard_1, come mostrato in figura 4.3.

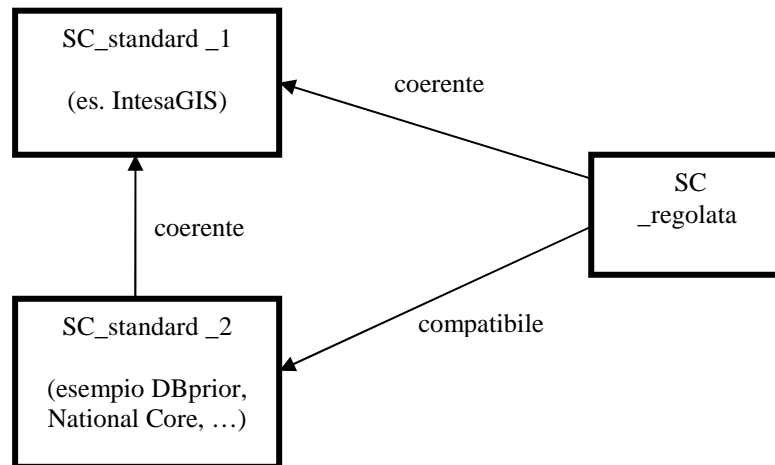


Figura 4.3

4.5 Livelli ulteriori di flessibilità

E' probabile che il livello di flessibilità raggiunto con le relazioni implementate in GeoUML catalogue 2008 non sia ancora sufficiente per trattare tutti i tipi di situazioni che si vorrebbero gestire, perché si tratta di una flessibilità basata in sostanza solo sui concetti di **selezione e contenimento** tra Elementi Informativi; in particolare questi concetti possono essere sufficienti nella definizione di nuove SC_regolate, ma risultare inadeguati per recuperare nel meccanismo di regolazione specifiche precedenti l'entrata in funzione delle regole.

Livelli più forti di flessibilità si possono sicuramente aggiungere in futuro utilizzando anche il concetto di **trasformazione** degli Elementi Informativi tra diverse specifiche (le trasformazioni possono riguardare sia la semplice ridenominazione di elementi tra specifiche diverse o anche trasformazioni più complesse che dovranno essere applicate ai dati). Alcune trasformazioni più semplici possono essere implementate da subito, altre, più complesse, richiedono un'attività di studio per mettere a fuoco le esigenze effettive che si vuole e si riesce a soddisfare.

5 I MODELLI IMPLEMENTATIVI E LA PRODUZIONE DI SCHEMI FISICI

5.1 Funzione del Modello Implementativo

Una Specifica di Contenuto a livello concettuale è volutamente astratta, nel senso che non definisce gli aspetti che dipendono dalla tecnologia. Per poter creare effettivamente un Dataset conforme a una SC è quindi necessario indicare le modalità di implementazione, che dipendono a loro volta dalla tecnologia adottata.

La definizione di tutti gli aspetti che permettono di realizzare una materializzazione effettiva di un Dataset costituisce il suo **Schema Fisico**. Nel contesto della GeoUML methodology gli schemi fisici devono essere prodotti automaticamente dalla Specifica di Contenuto in base a un insieme di regole dette **Modello Implementativo (MI)**.

E' importante osservare che il Modello Implementativo è costituito da un insieme di regole che non dipendono dalla SC alla quale viene applicato; in sostanza valgono ambedue le seguenti affermazioni:

1. lo stesso Modello Implementativo può essere applicato a diverse Specifiche di Contenuto, producendo diversi Schemi Fisici che condividono gli aspetti legati alla tecnologia, ma si riferiscono a contenuti diversi
2. la stessa Specifica di Contenuto può essere materializzata secondo diversi Modelli Implementativi, ottenendo schemi fisici che rappresentano in maniera diversa gli stessi contenuti.

Il modo in cui un modello implementativo è definito varia in base alla tecnologia di riferimento, in particolare in base alla modalità di rappresentazione dello schema fisico previsto dalla tecnologia. Ad esempio, per la tecnologia GML lo schema fisico è costituito da un XSD (XML Schema) e il Modello Implementativo consiste in regole che permettono di produrre tale XSD per una generica SC; per la tecnologia GeoRelazionale lo schema fisico è costituito da definizioni in SQL, quindi il modello implementativo deve supportare la generazione di quel tipo di definizioni.

In linea di principio il Modello Implementativo può essere fisso oppure parametrico:

- **MI fisso** è costituito da regole di implementazione che vengono applicate a tutti i contenuti senza la possibilità di indicare opzioni diverse sui singoli contenuti informativi
- **MI parametrico** contiene anche opzioni che devono essere specificate per ogni contenuto informativo

Esempio: si considerino gli attributi a tratti, cioè gli attributi il cui valore può variare lungo una geometria lineare; l'implementazione degli attributi a tratti può essere strutturale (cioè basata sulla creazione di geometrie che corrispondono al singolo tratto) oppure dinamica, basata sulla nozione di ascissa curvilinea. In un MI parametrico la regola per decidere quale implementazione adottare per un dato attributo a tratti deve essere indicata esplicitamente, mentre in un MI fisso è derivata automaticamente (nel caso più semplice è unica).

Il GeoUML catalogue supporta i modelli implementativi sotto i seguenti aspetti:

1. la struttura del formato di rappresentazione di una specifica (e corrispondentemente la struttura dati interna) prevede la rappresentazione di alcune informazioni necessarie per definire diversi modelli implementativi
2. la struttura del software prevede la possibilità di inserire diversi plug-in specializzati per produrre schemi fisici in base a diversi modelli implementativi

Inoltre nella versione 2008 è incluso un plug-in capace di produrre sia lo schema GML, sia gli XMI di diversi livelli di Application Schema, descritti nella seguente sezione. La definizione dettagliata delle regole adottate da questa funzione, cioè il Modello Implementativo adottato, è contenuta nel documento **“Generazione di Application Schema e di GML Schema da una Specifica GeoUML”**

5.2 Generazione di un Application Schema di alto livello

La comprensione approfondita di questo paragrafo e del successivo richiede la conoscenza di alcune nozioni relative alle tecnologie utilizzate nel contesto degli standard ISO TC211, in particolare il funzionamento fondamentale di uno strumento CASE per UML e il meccanismo di scambio di modelli tra strumenti basato sullo standard XMI.

GeoUML catalogue permette di derivare da una Specifica SC in GeoUML un Application Schema (AS) conforme agli standard ISO. Questo Application Schema rappresenta esattamente gli stessi contenuti presenti nella SC in GeoUML, con le seguenti precisazioni:

1. L'Application Schema è più complicato della SC, proprio perché GeoUML è stato progettato per semplificare la rappresentazione di certi contenuti
2. I vincoli GeoUML sono rappresentati nel AS come puri commenti associati alla classe vincolata e formulati secondo la sintassi GeoUML

La definizione esatta delle trasformazioni adottate per produrre l'AS è definita nel documento citato.

Concretamente l'AS viene esportato dal GeoUML catalogue in formato XMI, che è uno standard per scambiare modelli UML tra diversi CASE, e può essere importato da qualsiasi CASE (fatti salvi i problemi non banali di conformità delle versioni di UML ed XMI supportate dai diversi strumenti!). Dato che XMI standardizza solamente la parte relativa ai contenuti del modello (classi, attributi, associazioni, ecc...), ma non la relativa rappresentazione grafica, importando l'AS in uno strumento CASE è generalmente necessario sistemare la strutturazione grafica – ovviamente la complessità di questa operazione dipende dal tipo di strumento.

In figura 5.1 è rappresentata schematicamente questa forma di collaborazione tra GeoUML catalogue e un CASE, che permette di produrre un AS standard in forma di diagrammi UML (che possono essere ricaricati nel GeoUML catalogue come immagini di documentazione).

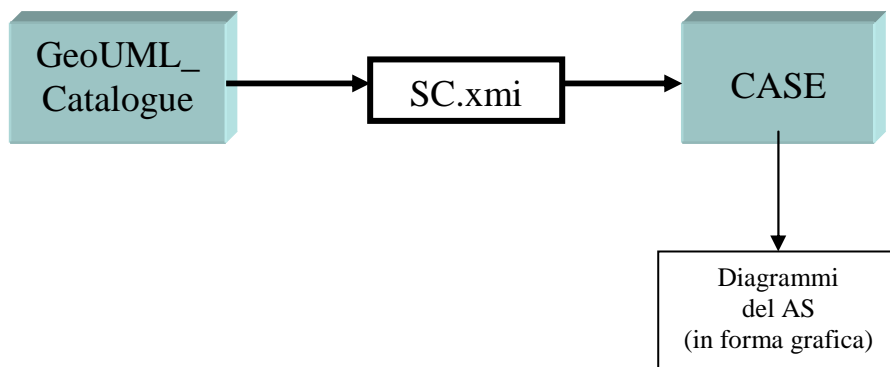


Figura 5.1

5.3 Generazione dello Schema GML_ESF di una SC

Lo standard ISO 19136 definisce nell'Annex E le regole per derivare da un AS un corrispondente XSD per GML. Sostanzialmente tali regole prevedono un processo in 2 passi:

1. Modificare l'AS introducendo una serie di elementi orientati alla produzione dello XSD per GML; chiamiamo "AS_{gml_oriented}" l'AS modificato secondo queste regole
2. Generare lo XSD per GML secondo regole abbastanza lineari.

Il GeoUML catalogue produce sia l'AS_{gml_oriented}, in formato XMI, che lo XSD per GML; l'AS_{gml_oriented} può essere caricato in un CASE, producendo i relativi diagrammi che possono essere utilizzati come documentazione dell'intero processo. Le regole dettagliate applicate per svolgere questa funzione sono definite nel documento citato.

Dato che esistono strumenti in grado di produrre uno XSD per GML da un l'AS_{gml_oriented}, in linea di principio è possibile produrre lo XSD per GML di una specifica SC seguendo 3 diversi percorsi, mostrati in figura 5.2, anche se solo la prima è raccomandata nella GeoUML methodology:

1. Generazione del XSD per GML direttamente dal GeoUML catalogue
2. Generazione dello XMI del l'AS_{gml_oriented} direttamente dal GeoUML catalogue e generazione del XSD per GML tramite uno strumento di codifica standard alimentato con tale XMI
3. Generazione dello XMI del l'AS_{gml_oriented} direttamente dal GeoUML catalogue e caricamento in un CASE tool, dove può essere editato; la generazione del XSD per GML in questo caso avviene tramite uno strumento di codifica standard alimentato dal XMI prodotto dal CASE tool

La terza modalità viene assolutamente sconsigliata, perché permette di basare la generazione del XSD per GML su un AS modificato nel CASE tool e quindi non garantisce l'allineamento tra XSD per GML e Specifica SC.

La seconda modalità produce uno XSD per GML sostanzialmente equivalente, ma più generico della prima.

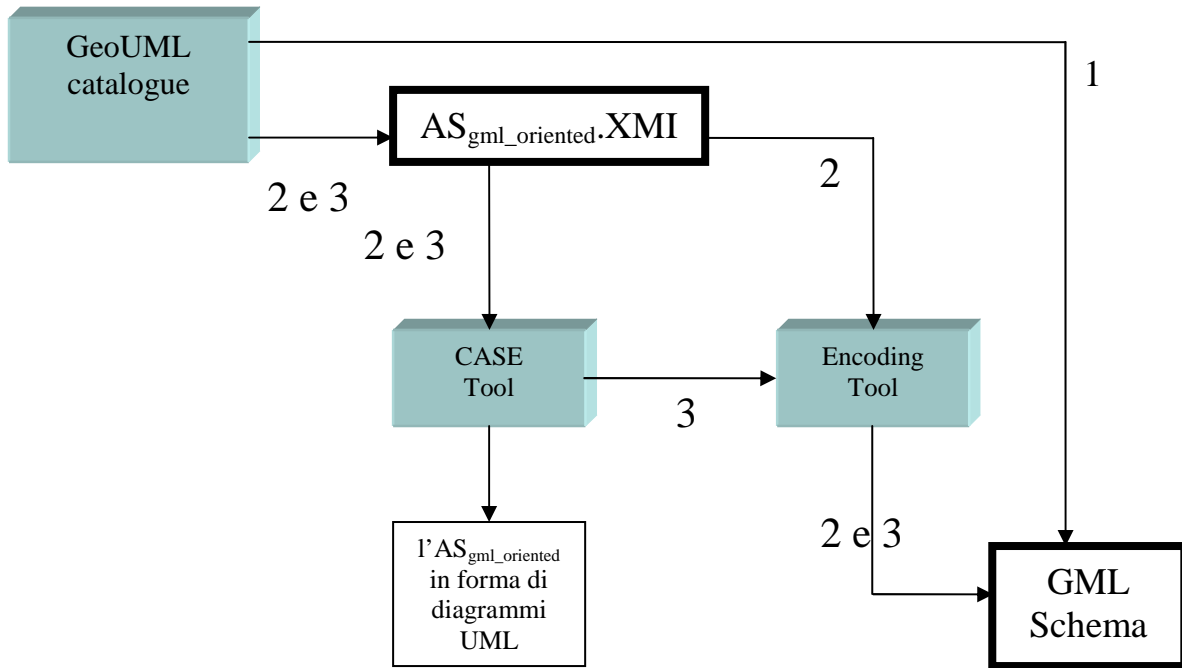


Figura 5.2 – Tre diverse modalità di generazione del XSD per GML; solo la prima è raccomandata

5.4 Altre Categorie di Modelli Implementativi

E' possibile estendere GeoUML catalogue con dei plug-in per produrre schemi fisici secondo altri modelli implementativi. Il costo di questa operazione è modesto e sicuramente il risultato presenta notevoli vantaggi rispetto alla definizione manuale di uno schema fisico.

Formati di scambio

Per quanto riguarda i formati di scambio, è opportuno mantenere molto limitato il numero di alternative, per non rendere l'interoperabilità eccessivamente complicata da gestire.

Sulla tecnologia GML il GeoUML catalogue già supporta il formato GML_ESF; nel documento generale è definita anche la struttura generale di un formato basato sul modello topologico (complessi).

Se non si vuole abbandonare la tecnologia tradizionale degli ShapeFile, potrebbe essere utile definire un formato adeguato a rappresentare tutti i contenuti di una specifica in ShapeFile+Tabelle secondo regole condivise.

Strutture di Database

Per i database GeoRelazionali potrebbe essere valido definire due Modelli Implementativi di riferimento, uno normale e uno topologico. Tuttavia uno strumento di generazione di tali schemi dovrebbe permettere alternative nella scelta delle strutture fisiche per permettere di ottimizzarle in base all'uso previsto, ed è quindi più complesso da progettare.

6. STRUMENTI SOFTWARE E MODELLO IMPLEMENTATIVO

La creazione effettiva di un Dataset conforme a una SC richiede nel concreto che sia definito anche un Modello Implementativo; quindi un Dataset oltre ad essere conforme a una SC deve sempre essere associato anche a un Modello Implementativo.

Qualsiasi Componente Software che voglia operare su un Dataset deve conoscere non solo la SC alla quale il Dataset è conforme, ma anche il Modello Implementativo utilizzato per materializzarlo.

Per evitare che ogni Componente Software debba essere realizzato in tante versioni quanti sono i Modelli Implementativi da supportare è possibile isolare le dipendenze dal Modello Implementativo in opportuni “Driver” relativi ai singoli Modelli Implementativi, capaci di importare ed esportare un Dataset in un modello interno del Componente Software . Ad esempio, in figura 6.1.a si mostra un generico componente software che legge un Dataset conforme a una SC e a un Modello Implementativo MIx; il componente utilizza un Driver per interpretare il modello implementativo MIx, ma la parte sostanziale del componente software dipende esclusivamente dalla SC.

Anche il GeoUML validator presentato in figura 3.1, pur essendo capace di operare su un Dataset rispetto a qualsiasi SC, deve possedere il Driver relativo al Modello Implementativo adottato per creare tale Dataset. (figura 6.1.b) .

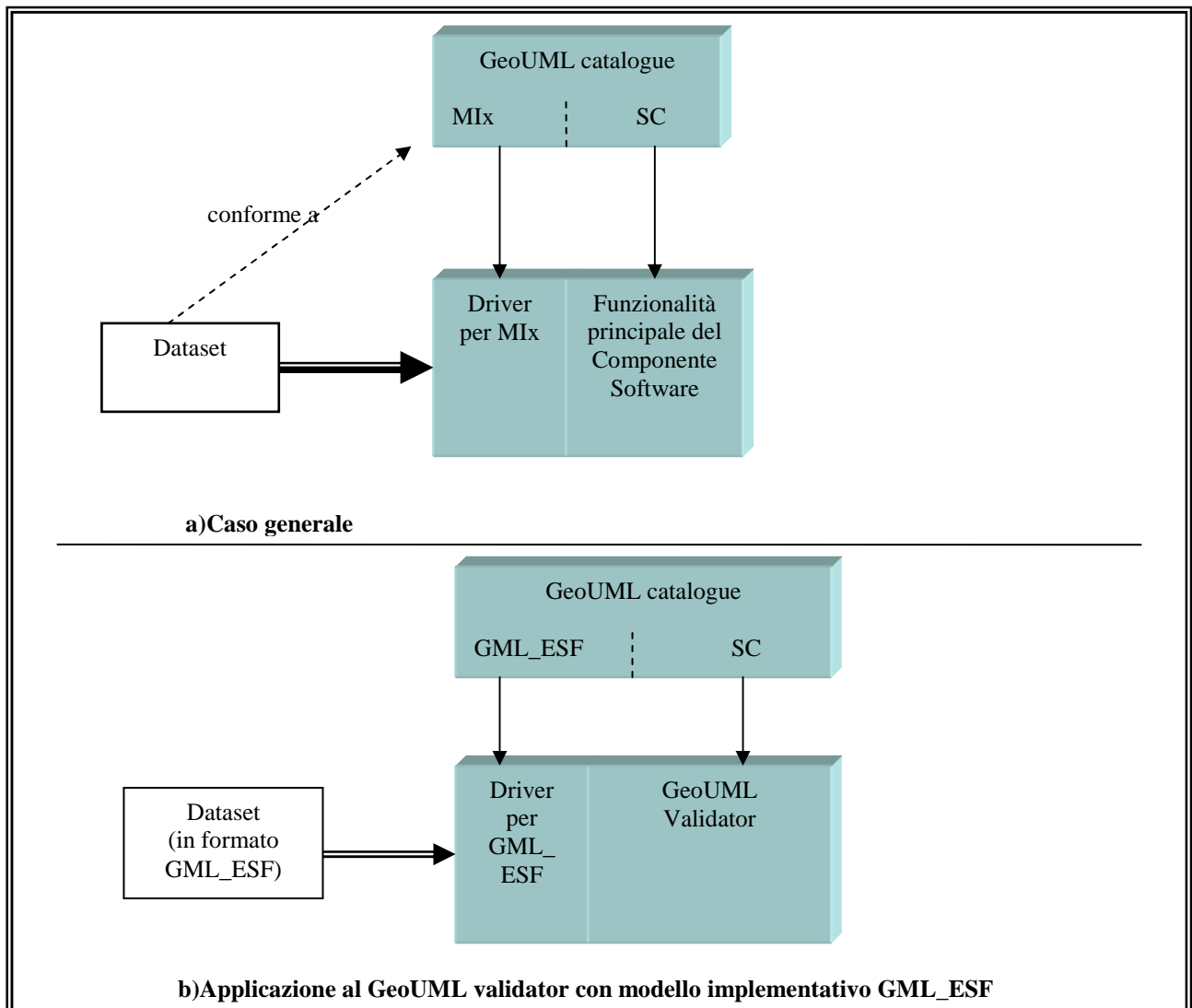


Figura 6.1 – Isolamento della dipendenza di un componente Software dal Modello Implementativo

Utilizzando questa impostazione, uno strumento per effettuare la conversione di un Dataset da un Modello Implementativo a un altro è realizzabile facilmente, se esistono i relativi driver (figura 6.2). In generale, la definizione di un nuovo Modello Implementativo dovrebbe essere accompagnata dalla realizzazione dei relativi driver.

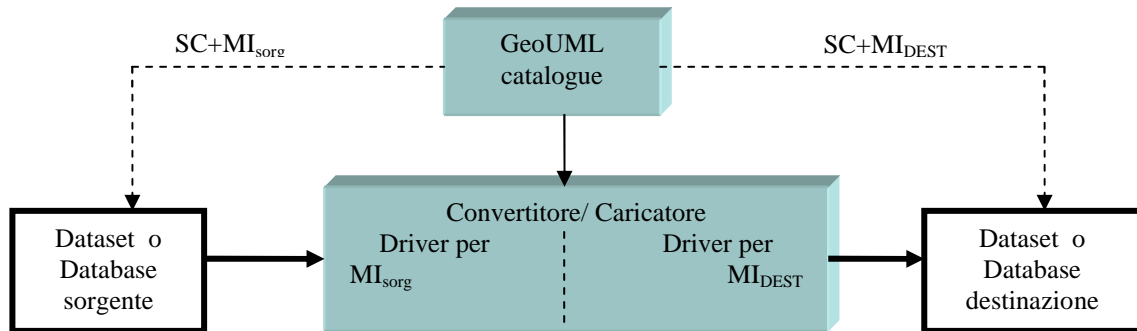


Figura 6.2

Appendice A1 – RAPPORTO TRA GEOUML METHODOLOGY E GLI STANDARD ISO TC211

Impostazione generale

La GeoUML methodology copre un'area di problemi già regolamentati dagli standard ISO TC211.

Alla base della GeoUML methodology è posto il principio di essere compatibile con l'impostazione ISO TC211 in modo da poter usufruire di tutta la strumentazione conforme a tali standard che verrà presumibilmente prodotta; tuttavia la GeoUML methodology estende l'impostazione del ISO TC211 per venire incontro a tre tipi di esigenze:

1. aumentare, rispetto a tali standard, il **disaccoppiamento tra le Specifiche di Contenuto e le Strutture Implementative**, in modo da rendere il livello di astrazione delle Specifiche di Contenuto più conforme alla nozione classica di un Modello Concettuale, che è quella di una "descrizione dei contenuti informativi precisa ma indipendente da tutti gli aspetti di implementazione (che dipendono dalla tecnologia)";
2. **gestire con un unico strumento** non solo la documentazione strutturata, ma anche la documentazione descrittiva, destinata ad essere letta dagli esseri umani, garantendo in tal modo l'allineamento tra questi due tipi di documentazione;
3. **gestire le relazioni di coerenza e compatibilità** tra specifiche diversi.

Si analizzano ora le motivazioni delle prime due esigenze di arricchimento dell'impostazione ISO enunciate sopra e della soluzione proposta nella GeoUML methodology; per quanto riguarda la terza esigenza, non esiste al momento nell'impostazione ISO alcun riferimento a questo tipo di problema.

Disaccoppiamento tra le Specifiche di Contenuto e le Strutture Implementative

Nell'impostazione ISO esistono due componenti per la definizione dei contenuti di una specifica: l'Application Schema (AS) e il Feature Catalogue (FC).

Nell'impostazione ISO TC211 il livello di astrazione del AS (e, in parte, anche del FC) è più dipendente dalla implementazione di quanto non sia opportuno per una Specifica di Contenuto concettuale; in particolare gli aspetti che nella GeoUML methodology si è ritenuto di dover correggere sono i seguenti:

- a. i tipi geometrici ISO definiscono tutti gli aspetti di implementazione (interpolazione ecc...) che sono irrilevanti nella analisi e comprensione dei contenuti (ad esempio, a livello concettuale non è necessario distinguere se una "curva" è implementata come linestring, curva composta, ecc...), sono sostituiti nella GeoUML methodology da tipi più generici che vengono poi trasformati in tipi ISO tramite il Modello di Implementazione
- b. a causa delle regole molto strette per la produzione del GML definite in ISO 19136 l'AS deve contenere scelte di dettaglio che determinano il tipo di codifica GML che viene prodotto, ma che sono irrilevanti a livello concettuale e incomprensibili ai non esperti di GML
- c. la specifica dei vincoli in ISO TC211 è carente – il problema viene liquidato facendo riferimento al linguaggio OCL, che però non è adatto allo scopo per diversi motivi:
 - c1. OCL non è integrato con gli aspetti relativi ai dati geometrici, le relazioni topologiche, ecc..
 - c2. OCL non è sintatticamente adatto a una specifica concettuale, essendo eccessivamente difficile per un utente non specialista
 - c3. OCL non è implementabile
- d. mancano in ISO alcuni costrutti ad alto livello, come gli Attributi a Trattati e gli Attributi a Sottoaree, che devono essere rappresentati tramite la loro struttura implementativa in termini di Geometrie Composte (in questo caso risentendo oltretutto della carenza nell'espressione dei vincoli citata al punto precedente) oppure in forma di ascissa curvilinea, per la quale non è ancora definito l'apposito standard; nella GeoUML methodology la distinzione tra i diversi tipi di implementazione degli Attributi a Trattati è confinata nel Modello di Implementazione
- e. mancano in ISO TC211 costrutti ad alto livello per gestire la descrizione delle geometrie 2D e 3D, che caratterizzano i contenuti che vengono realizzati allo stato attuale dell'arte, caratterizzato da una impossibilità di realizzare un modello tridimensionale completo.

Avere isolato nella nozione di Modello di Implementazione gli aspetti che non appartengono al livello concettuale permette all'impostazione della GeoUML methodology di ottenere in ultima analisi due benefici:

1. la definizione delle Specifiche di Contenuto, che costituisce un processo complicato di analisi applicativa e confronto tra punti di vista disciplinari diversi, può concentrarsi solamente sugli aspetti veramente di contenuto, senza dover tener conto di aspetti implementativi, e permettere il confronto tra agli esperti applicativi senza la mediazione degli specialisti di tecnologia
2. le Specifiche di Contenuto non sono impattate da ogni evoluzione della tecnologia geomatica e possono resistere più a lungo nel tempo; è infatti possibile produrre nuovi Modelli Implementativi per supportare tale evoluzione, salvaguardando l'investimento in analisi applicative.

Gestione delle Specifiche con uno strumento unico

La strumentazione disponibile per gestire l'AS e il FC definiti da ISO non è unica; in particolare nel modello di gestione delle specifiche del ISO TC211 il componente fondamentale è il generico strumento CASE UML che gestisce l'AS. In effetti, questa scelta è comprensibile alla luce della storia dello sviluppo delle specifiche; ad esempio, tutto lo Spatial Model non avrebbe potuto essere facilmente prodotto senza questo approccio.

Tuttavia, nelle specifiche di contenuto l'esistenza dei diversi Package Geometrici è irrilevante; inoltre, nei database territoriali la scarsa numerosità delle associazioni presenti rispetto al numero di classi e di attributi rende l'uso di diagrammi UML assai poco utile. In sostanza, rispetto alla funzionalità di uno strumento CASE UML, le funzionalità veramente utili alla rappresentazione di una specifica di contenuto sono relativamente poche, mentre mancano le funzionalità necessarie alla gestione dei numerosi e importanti aspetti descrittivi.

Per rispondere a questo difetto, nell'impostazione ISO TC211 l'AS gestito dal CASE UML è affiancato da un Feature Catalogue che deve essere gestito con un diverso strumento. Il rapporto tra queste due rappresentazioni dei contenuti non è chiaro, e in effetti diversi enti in nazioni diverse hanno proceduto in maniera diversa (produzione prima del AS e poi del FC o viceversa).

Il GeoUML catalogue si propone di superare queste limitazioni tramite l'inserimento a monte rispetto ad AS e FC di una documentazione integrata rigorosamente di livello concettuale; da questa documentazione è possibile produrre l'Application Schema e il Feature Catalogue previsti da ISO TC211. Tale produzione avviene tramite la definizione di un Modello di Implementazione che aggiunge alla Specifica di Contenuto la definizione delle scelte implementative.

Nell'impostazione della GeoUML methodology il GeoUML catalogue è più simile a un Feature Catalogue, ma contiene molti elementi di un AS, con le precisazioni relative al livello di astrazione svolte nella sezione precedente, superando in tal modo i limiti degli strumenti CASE.

Appendice A2 – SPECIFICHE INTESAGIS E GEOUML METHODOLOGY

Le Specifiche IntesaGIS coprivano sia aspetti di contenuto, sia aspetti legati alle regole generali utilizzate per gestire tali contenuti.

Sostanzialmente la GeoUML methodology incorpora dalle specifiche IntesaGIS tutti gli aspetti relativi alle regole generali, lasciando intatte le specifiche di contenuto (documenti IntesaGIS 1007_1/2 e 1007_4).

In questo modo la GeoUML methodology definisce una tecnologia per la gestione delle specifiche di un database territoriale che è più generale rispetto agli specifici contenuti IntesaGIS.

L'esigenza di attuare questa distinzione netta è emersa chiaramente già per produrre lo Schema di codifica in GML delle specifiche IntesaGIS di contenuto; non aveva infatti nessun senso codificare un tantum tale Schema GML utilizzando un editor XML, come se si trattasse di una definizione di un nuovo contenuto indipendente, piuttosto che implementare un codificatore automatico capace di generare lo Schema GML direttamente dalle specifiche.

In termini di documentazione questo approccio coincide con la sostituzione del documento di contenuto previsto da IntesaGIS (1007_5 – codifica in GML) con delle regole tecniche generali per la produzione dello Schema GML .

Per quanto riguarda il ruolo che le specifiche di contenuto IntesaGIS devono svolgere, nella GeoUML_Methodology si possono fare solamente le seguenti affermazioni:

- le specifiche di contenuto IntesaGIS possono essere caricate nel GeoUML catalogue (in effetti sono state già caricate) e i documenti 1007_1/2 e 1007_4 possono essere prodotti in forma quasi identica da tale strumento
- la possibilità di definire diverse specifiche di contenuto e di regolamentare le relazioni tra di esse possono contribuire a precisare e differenziare il ruolo delle specifiche IntesaGIS

Come utilizzare queste possibilità è materia di scelta dei contenuti ed esula dagli obiettivi della GeoUML methodology.