

## Il Multi-database ed il mito ultimo dell'interoperabilità semantica

Claudia Pegoraro <sup>(a)</sup>

<sup>(a)</sup> Professionista indipendente, via Richelmy 25/A 10126 Torino, 335 1395923, [claudia\\_pegoraro@yahoo.it](mailto:claudia_pegoraro@yahoo.it)

### Riassunto

Quando nei primi anni duemila i documenti guida della Direttiva INSPIRE lanciarono il *multi-database* come mediatore “leggero” dell'interoperabilità semantica, a molti parve finalmente tracciata la via risolutiva per addivenire all'obiettivo, a beneficio delle innumerevoli fonti di dati geografici o “spaziali” (nell'accezione più internazionale del termine) disseminate nel continente europeo. “Leggero” perché inteso come poco invasivo e quindi rassicurante, in quanto strumento abile nel garantire un risultato universalmente atteso e di riconosciuta utilità senza distruggere o invalidare l'esistente.

In questa dissertazione si ripercorreranno in veste critica i principali passi del processo di allestimento e messa in campo del *multi-database* INSPIRE, ovvero: la fase di definizione metodologica, la fase di modellazione concettuale, la fase di recepimento normativo da parte degli Stati Membri, la fase di implementazione. L'intenzione è quella di soffermarsi dapprima sugli obiettivi attesi/disattesi. Secondo aspetto di approfondimento: le difficoltà e gli ostacoli incontrati soprattutto durante la fase di implementazione, differenziando quelli di interesse generalizzato per tutti gli Stati Membri da quelli che maggiormente contraddistinguono la situazione italiana. Terzo e ben più ambizioso ambito di argomentazione quello inerente le possibili strategie correttive applicabili “in corsa” nell'attuale fase di implementazione, ovvero le buone pratiche per un efficace lavoro di trasformazione semantica finalizzato all'interrelazione delle specifiche dati nazionali, in particolare quelle italiane, con le specifiche dati comunitarie. È ormai indubbio che trascurando specifici ingredienti, dosati nell'ambito di un approccio metodico e sistematico, non si possa addivenire ad un *repository* organico ed unitario di dati anche solo a livello nazionale. E soprattutto non è possibile trascurare l'obiettivo finale e più pragmatico, cioè quello di ottenere con una certa “speditezza” una cartografia multi-tematica omogenea ed armonizzata su di un determinato ambito territoriale, nazionale o comunitario, che sia ospitata all'interno di una *Spatial Data Infrastructure* e sia in grado di supportare opportunamente le necessarie scelte politico-amministrative: che, a ben vedere, era all'origine uno dei *focus* primari della stessa Direttiva INSPIRE.

### Abstract

In 2002 the position papers of the INSPIRE Directive proposed the *multi-database* as the perfect solution for semantic interoperability purposes, to reach finally a real harmonization for spatial sets among several data sources across Europe. This as a “light” approach to grant long overdue results with no remake of existent data archives.

Here we will propose a critical review of a standard process towards the *multi-database* objective, taking the INSPIRE one as the pillar-reference. All the steps including the methodological phase, conceptual modelling, Member States adoption and implementation phase will be considered to look at objectives and troubles along the way. In the actual phase, the implementation phase, this review will be strengthened and focused more on the Italian context.

Finally some repair and optimization strategies will be produced, coming from best practice experience, to get real synergy and interoperability for data specifications, considering the national level and the European level to be harmonized. Moreover those best practices are reusable at different administrative levels, that is to get interoperability as a national repository, when the

national level and the local levels need to be harmonized. And the same best practices are able to promote the last objective, and the most important and operational one declared by INSPIRE: to get finally a high quality multi-thematic cartography extended to wide territorial areas, this one to reach a really harmonized Spatial Data Infrastructure and to encourage spatial information as a follow-up to the implementation of major European environmental policies.

### **Introduzione**

Il *multi-database* è tutt'altro che superato. Semmai non è stato ancora raggiunto e neppure così compreso.

I processi di integrazione tra dati geografici ed alfanumerici puri si sono moltiplicati e l'interoperabilità tra sistemi informativi ha fatto, con grande sforzo, più di un passo in avanti, tuttavia diverse delle potenzialità del *multi-database*, soprattutto alcune delle più promettenti, non sono state neppure sfiorate e rimangono lontane dal manifestarsi. Molti interventi di modellazione sono stati intrapresi in quest'ottica ed ad ampio raggio territoriale, nazionale ed anche continentale, ma alcuni appaiono a tratti "statici", poco versatili, se non addirittura poco operativi.

Emblematico in tal senso il caso INSPIRE, che verrà ripercorso nelle sue fasi principali prendendone ad esempio i punti di forza e le criticità. Ma partiamo direttamente dalla sequenza degli *step* operativi che conducono a costruire un modello dati o "modello semantico" e che sono stati affrontati anche nel caso della Direttiva.

### **Il processo INSPIRE e la modellazione semantica**

INSPIRE si appresta dapprima, a partire dal 2005 e con la costituzione dei *Drafting Team*, a definire accuratamente la parte metodologica appoggiandosi alle canoniche modalità di *Unified Process* ed alla formalizzazione in linguaggio *UML*, affiancate dall'adozione degli standard internazionali della serie *ISO 19100 Geomatics* e cercando di costruire un "manuale" per i modellatori dell'informazione, che verranno successivamente impegnati sui trentaquattro temi della Direttiva. Questa ragguardevole mole di regole e di substrati culturali costituisce la base portante sulla quale i modellatori devono poi, con non poco sforzo, operare: la fase di modellazione concettuale dell'informazione parte dalla raccolta dei requisiti, attività secondo la quale il contenuto di natura tematica deve essere raccolto con modalità "tecnica", quasi maieutica. Ma, come troppo spesso avviene, si presenta il rischio di uno sbilanciamento dei ruoli, con gli esperti tematici che si trovano a ricoprire sì il ruolo di "consulenti", ma anche di decisori sulla formalizzazione del modello concettuale al posto degli analisti, nella prima forma del modello ed ancor più nelle sue successive evoluzioni e revisioni; mentre questo secondo ruolo spetta di necessità ad un analista esperto.

Nella fase dedicata alla formalizzazione del modello concettuale altrettanto facilmente si presta attenzione al "tema", ma è diffusa la mancanza di attenzione agli oggetti afferenti a tematiche trasversali, che andrebbero di regola considerati in un'ottica aperta ai più disparati contesti tematici, in cui potrebbero trovarsi, anche solo potenzialmente, coinvolti (Batini, Lenzerini, 1984).

Questa sola è la modalità vincente per modellare l'informazione in maniera da favorire l'interoperabilità semantica, lavorando con profondità di dettaglio sullo specifico contesto tematico, ma con un'apertura "enciclopedica" ai concetti non strettamente afferenti al tema, la cui connotazione andrà ricercata in un ambito multi-tematico e multi-culturale (Batini et al., 1986).

Attenzione va riservata anche ai differenti livelli di "profondità amministrativa" della modellazione: è indubbio che un requisito recepito al livello nazionale andrà opportunamente "tradotto ed universalizzato" per passarlo al livello comunitario, così come pure esiste differenza tra un archivio al livello comunale e il medesimo archivio al livello nazionale: eppure ospitano pressoché il medesimo contenuto! Questo aspetto appare notevolmente contraddittorio e complesso da illustrare, poiché richiede davvero grande esperienza e volontà di approfondimento da parte dell'analista che si appresta a progettare il modello. Purtroppo è una delle chiavi di interpretazione fondamentali, seppur tra le più trascurate, dell'interoperabilità semantica: questo perché il contenuto informativo

di due archivi può essere apparentemente coincidente, ma diversa l'esigenza ed il processo grazie ai quali sono nati e diversa quantomeno la profondità di taluni dettagli, tra i quali i metadati e gli eventi operazionali a corredo.

In INSPIRE spesso le esigenze manifestate dagli esperti tematici riguardano la riproposizione più fedele possibile dei propri archivi di riferimento, mentre il *focus* o confine del modello viene perso di vista: il “taglio” e l'insieme di informazioni da ricomprendere nelle specifiche dovrebbe rispettare l'obiettivo primario di materiale atto ad assecondare le diverse politiche comunitarie di carattere ambientale e non ad alimentare il *repository* teoricamente onnicomprensivo di tutti i dati esistenti a tutti i livelli amministrativi. Esso si rivela invece destinato a mantenersi obiettivamente generico per via dei conflitti esistenti assolutamente non sanabili se mantenuti su di un piano “di parte” piuttosto che “universalizzati” (Lillethun et al., 2002).

L'armonizzazione di modelli concettuali esistenti, nel caso di INSPIRE quelli facenti capo ai trentaquattro temi della Direttiva, si impone come ulteriore tassello cruciale per il processo (Rase et al., 2002): si lavora cercando di sanare gli inevitabili conflitti, sovrapposizioni e ridondanze causati dall'operare in parallelo di trentaquattro distinti modellatori. Questo tentativo va solo parzialmente in porto, per varie ragioni strutturali: l'esperienza insegna che la supervisione di tante attività di modellazione, volta alla conciliazione di ridondanze ed incoerenze, è una responsabilità che andrebbe fortemente accentrata e demandata ad un unico decisore tecnicamente capace e *super partes*, piuttosto che alla strategia di parlamentare, seppure sottoposta ad arbitrato, tra gruppi eterogenei (Zhang et al., 2002).

### **Le problematiche dell'attuale fase di implementazione**

INSPIRE affronta successivamente la fase di recepimento normativo e di implementazione ai vari livelli nazionali, intesa come la messa in atto di strategie tecnologiche concepite quali “ponti di collegamento” tra la Direttiva ed i fronti nazionali interni, ai fini dell'implementazione delle specifiche tecniche comunitarie. Si lavora parecchio al modello di interscambio dati in linguaggio GML e ad approntare gli strumenti e le strategie di trasformazione dei dati (si vedano il progetto HALE e le diverse iniziative nazionali).

Ancora una volta però le difficoltà non sono prevalentemente tecnologiche, ma semantiche, per ciascuno degli Stati membri; facciamo soltanto breve cenno alle principali: il *gap* esistente tra specifiche dati, metadati e servizi web per l'informazione geografica; le specifiche comunitarie che risultano troppo poco armonizzate tra i vari tematismi; il *gap* semantico tra le specifiche comunitarie e le specifiche dati nazionali, che dovrebbe condurre ad un certo ripensamento ed ottimizzazione delle seconde nell'ottica di avvicinamento alle esigenze transnazionali.

A questi aspetti si aggiungono l'obiettiva complessità di trattamento dei dati, spaziali e non, provenienti da diverse fonti (si pensi soltanto alla mancata corrispondenza delle forme geometriche sul limite dei confini di Stato o semplicemente sul limite territoriale di competenza di enti amministrativi distinti). Ancora la disponibilità non immediata dei dati (la maggioranza delle fonti, in Italia soprattutto, non è ancora *open*) e la loro frammentarietà, duplicazione, ridondanza e non coerenza rendono abissali le tempistiche previste per la razionalizzazione dei contenuti informativi e, di fatto, la tanto auspicata copertura cartografica multi-tematica unitaria sul territorio continentale, da porre a supporto delle politiche comuni, si rivela un progetto appena abbozzato.

### **Alcuni spunti per possibili strategie correttive**

Al termine di questa trattazione si ha l'ambizione di ipotizzare alcune strategie correttive “in corsa”, ovvero quali potrebbero costituire le caratteristiche di un buon lavoro di trasformazione semantica per favorire al massimo, allo stato attuale, il necessario allineamento tra le specifiche dei dati nazionali colle specifiche dei dati INSPIRE.

E qui la nostra disamina si sposta maggiormente sul fronte interno, concentrandosi sul caso italiano:

- I passo di allineamento, di natura strettamente semantica: sarebbe fondamentale “apparentare” il più possibile le due specifiche, recependo dalle specifiche comunitarie la parte migliore per costruire una “sovrastuttura” molto essenziale di alto livello alla struttura del modello dati nazionale, tema per tema. Si tratta di un lavoro di modellazione pensato nell’ottica di migliorare la mappatura dell’informazione sui due fronti, che altrimenti rimarrà costantemente frammentaria e poco organica.
- Il passo di allineamento, di natura semantica e tecnologica: sarebbe necessario riconvertire gli stereotipi del GeoUML, adottati in sede nazionale, negli originari stereotipi spaziali dello standard ISO della serie 19100, in quanto esattamente questi ultimi vengono utilizzati nel linguaggio di interscambio dell’informazione geografica GML, standard internazionale adottato anch’esso dalla Direttiva. Il GML è infatti il “veicolo” dell’informazione, la variante del linguaggio XML studiata appositamente per i dati geografici; è il linguaggio formale grazie al quale risulta codificato il modello di interscambio dei dati INSPIRE e dunque lo strumento che permette di dialogare con i servizi di interscambio ed acquisizione dei dati già approntati a livello centralizzato in sede europea.
- III passo di allineamento, di natura semantica: sarebbe necessario accelerare la procedura di “anti-frammentazione” dell’informazione, promuovendo la creazione di un’unica versione certificata per i principali *set* di dati, standardizzata ed unificata a livello nazionale, persistente nel tempo, ma ugualmente aggiornabile direttamente dai livelli amministrativi di competenza; ciò presuppone la creazione di un *repository* centralizzato multi-tematico dell’informazione molto ben strutturato ed articolato, non disgiunto da appositi processi per l’alimentazione e la gestione dell’informazione da parte degli attori competenti.
- IV passo di allineamento, strettamente tecnologico: sarebbe importante promuovere su larga scala sperimentazione ed impiego dei servizi di comparazione delle specifiche e di veicolazione dell’informazione già prototipati e messi a disposizione da parte dei Centri di ricerca della Commissione Europea e dai progetti comunitari finora avviati.

## Conclusioni

Trascurando un approccio di lavoro sistematico e metodologicamente unitario è difficile prevedere in quali tempi ed in quali modalità si potrà addivenire, quanto meno a livello nazionale, ad un *repository* centralizzato multi-tematico, che abbia l’ambizione di coprire, in maniera organica e funzionale, il più ampio e completo contenuto informativo dal punto di vista della componente alfanumerica integrata colla componente spaziale. Ci auguriamo che la prospettiva teorica e tecnologica portata in campo dal *multi-database* possa sempre più attuarsi ai diversi livelli amministrativi, nazionali e trans-nazionali, avvicinando finalmente il raggiungimento pieno dell’obiettivo interoperabilità per i sistemi che ospitano contenuti informativi di fondamentale interesse.

## Riferimenti bibliografici

- Batini C., Lenzerini M. (1984), "A Methodology for data schema integration in the Entity Relationship model", *IEEE Transactions on Software Engineering*
- Batini C., Lenzerini M., Navathe S.B. (1986), "A Comparative Review of methodologies for database integration", *IEEE Computing Surveys*
- Lillethun A., Weber J. L., Steenmans C. (2002), "Environmental Thematic User Needs - Position paper", *INSPIRE Environmental Thematic Coordination Position paper*, European Environmental Agency
- Rase D., Björnsson A., Probert M., Haupt M. (2002), "Reference Data and Metadata Position Paper", *INSPIRE Reference Data and Metadata Position paper*, EUROSTAT
- Zhang S., Wu X., Zhang C. (2003), "Multi-Database Mining", *IEEE Computational Intelligence Bulletin*, Vol. 2 N°1