

DBT e INSPIRE: un percorso semplificato per i dati

Gianni Campanile

Esri Italia, Via Tiburtina 755, 00159 Roma, Tel. 06 40696245, email: campanile@esriitalia.it

Riassunto

La direttiva INSPIRE è ormai da tempo nella fase di implementazione e le prossime scadenze riguardano l'implementazione di dati e servizi. Per quanto riguarda i dati, le specifiche sono ormai consolidate per i temi del primo allegato, e in corso di definizione e revisione per gli altri temi definiti negli allegati II e III. A breve quindi le pubbliche amministrazioni dovranno rendere disponibili i loro dati nella forma definita dalle specifiche INSPIRE. Spesso però è difficile convertire le basi dati esistenti, a causa di applicazione preesistenti, quindi è necessario mantenere la base dati originaria e definire un workflow di trasformazione che, attraverso procedure ETL (Extract-Transform-Load) mantenga allineate le due basi dati.

Per l'Italia però, non esiste solo la specifica INSPIRE ma anche la specifica dei database Geotopografici (DBT) che anch'essa definisce una struttura di base dati. Questa specifica, esistente da tempo ma divenuta legge a Febbraio 2012 non è direttamente compatibile con le specifiche INSPIRE; è quindi necessario per i soggetti interessati definire una seconda conversione dati, dal proprio sistema al formato DBT. La gestione di tutte queste trasformazioni, spesso complesse, e delle varie banche dati è sicuramente onerosa, sia in termini di infrastruttura che in termini di gestione. Oltre ad identificare con chiarezza il problema, Esri Italia sta definendo, per i temi compatibili, una trasformazione DBT-> INSPIRE tramite l'utilizzo di strumenti ETL sofisticati. In questo modo non è più necessario per l'utente definire due trasformazioni, ma solamente quella dal proprio sistema al DBT, essendo la seconda già presente.

Obiettivo di questo documento è quindi quello di mostrare i passaggi necessari e dei casi d'uso per portare dei dati strutturati in modo generico al mondo INSPIRE e DBT, dimostrando in pratica come sia possibile semplificare il problema pur aderendo alle specifiche dei dati e servizi europee ed italiane.

Abstract

INSPIRE Data specifications for Spatial data Themes of Annex I are now finalized, while many Annex II Themes specifications are already available even if in a draft stage. Thus Public Administrations must convert their data in the format defined by INSPIRE. Often this is not straightforward, and a better solution is to defined a transformation workflow that, using ETL (Extract-Transform-Load) procedures, keeps database up-to-date.

Italy has defined a different specifications for databases called DBT. These specs often are not directly compatible with INSPIRE specifications. Thus users should define an additional data transformation from their proprietary format to DBT. The management of these complex transformations and of all different data specifications is costly. Esri Italia, having clearly identified and studied this problem, is defining an ETL transformation from DBT to INSPIRE. This allows users to limit the effort on the single transformation from proprietary DB to DBT.

This paper is focused on steps needed to bring a proprietary DB to INSPIRE e DBT formats, showing in practice how this problem can be simplified still maintaining strict conformance to European and Italian standards.

Riferimenti:

1. INSPIRE Metadata: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/101>
2. INSPIRE Data specifications: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>
3. INSPIRE Network services: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/5>
4. Database geotopografici - DM 10 novembre 2011
 - a. http://www.digitpa.gov.it/sites/default/files/normativa/PCM_DBT_su_sito.pdf
 - b. http://www.digitpa.gov.it/sites/default/files/normativa/PCM_DBT_all_1_sito.pdf
 - c. http://www.digitpa.gov.it/sites/default/files/normativa/PCM_DBT_all_2_sito.pdf

Keywords:

Metadati, Dati, servizi, INSPIRE, standard, catalogo, geoportale, ETL, DBT.

1. Introduzione

La Direttiva INSPIRE, sempre più largamente conosciuta ed adottata, prevede delle specifiche ben precise per il formato dei dati per i suoi temi di riferimento. In realtà la Direttiva è un disegno complessivo, in cui dati e metadati sono raggiunti attraverso l'uso dei servizi definiti coadiuvati dalle specifiche e dalle raccomandazioni di interoperabilità (vedi *Figura 1*)

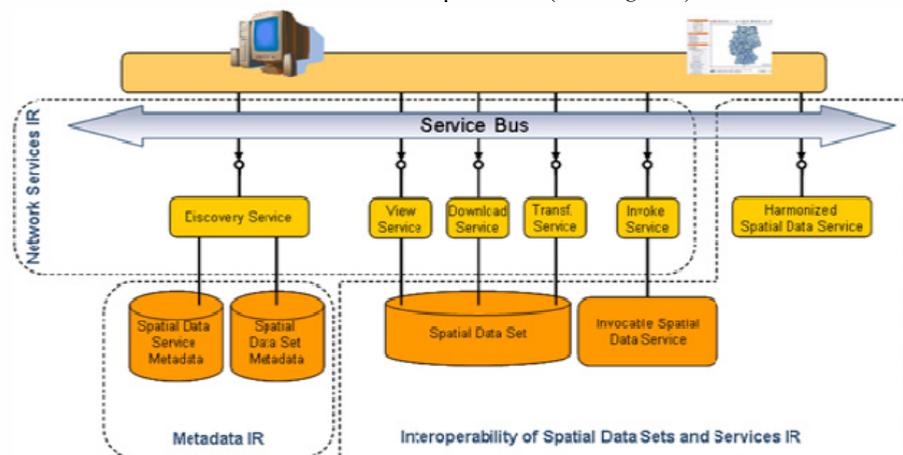


Figura 1. Architettura INSPIRE.

Tenendo comunque a mente il modello generale ma concentrandoci sui dati, dobbiamo prendere in considerazione le Data Specifications di INSPIRE (vedi [2]), un insieme di documenti corposi e complessi.

Pur essendo chiaro l'obiettivo finale però, dal punto di vista utente non è chiaro il *percorso migliore* per raggiungerlo; in altre parole: quali sono le azioni necessarie per avere i dati nel formato richiesto. L'esperienza fatta da Esri Italia con i diversi progetti anche a livello europeo ha permesso di definire alcuni prodotti ed un workflow di trasformazione che permette di rispondere a questa domanda, come già presentato ad ASITA 2011.

A Febbraio 2012 però, sono state approvate in GU le specifiche del database Geotopografico, che quindi impone necessità di soddisfare non solo le specifiche di INSPIRE ma anche quelle del DBT. Le trasformazioni necessarie diventano quindi due!

Nei capitoli che seguono il problema verrà dettagliato e verrà illustrata una soluzione per ridurre la complessità complessiva.

2. INSPIRE e DBT

Il problema della conversione dati si presenta per tutti quegli elementi che differiscono fra le specifiche INSPIRE e quelle del DBT. I motivi di queste differenze nascono sia dal fatto che, malgrado la recente approvazione, il DBT è stato definito diverso tempo prima che fossero disponibili le specifiche INSPIRE, sia dal fatto che INSPIRE si occupa di una realtà europea, e quindi nella definizione dei dati ha utilizzato spesso l'approccio del "minimo comun denominatore", cercando così di minimizzare l'impatto sui vari paesi e di non favorire un formato rispetto ad un altro.

Per chiarire questo aspetto, prendiamo in esame le unità amministrative di INSPIRE, cioè il Tema Administrative Units dell'allegato 1.

La Figura 2 mostra un dettaglio delle specifiche di implementazione.

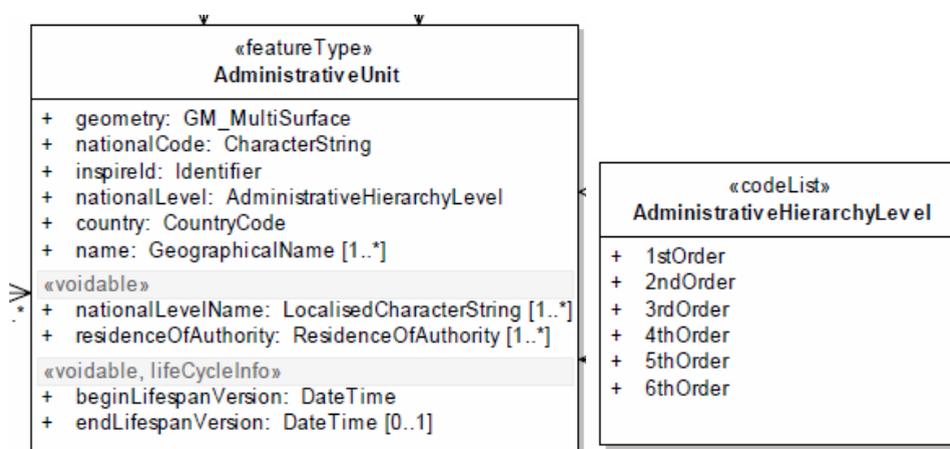


Figura 2. Particolare delle specifiche delle Administrative Units.

Come si può vedere le *AdministrativeUnit* sono rappresentate con dei poligoni e la loro classificazione è effettuata attraverso l'attributo *nationalLevel*, codificato con i valori gerarchici mostrati nella tabella *AdministrativeHierarchyLevel*.

In questa struttura, la suddivisione italiana in Stato, Regioni, Province e Comuni viene implementata creando 4 record:

Italia	<i>nationalLevel</i> = 1
Lazio (esempio di Regione)	<i>nationalLevel</i> = 2
Roma (esempio di Provincia)	<i>nationalLevel</i> = 3
Anzio (esempio di Comune)	<i>nationalLevel</i> = 4

E' stato stabilito che a livello europeo non ci sia (o quantomeno non interessi) una suddivisione oltre il sesto livello.

Questa struttura, se pure funzionante, non è sicuramente efficiente: ad esempio, per avere i comuni di una regione si deve effettuare una interrogazione gerarchica con un doppio join sulla stessa tabella, passando dalla regione alla provincia e poi al comune.

Questa gerarchia è invece ben esplicitata nello schema adottato dal DBT, che prevede una classe Regione, una Provincia e un Comune (non incluso in figura) legate da opportune relazioni:

CLASSE: Regione (REGION - 090106)	CLASSE: Provincia (PROVIN - 090105)
Popolamento della classe	Popolamento della classe
Definizione Ambito territoriale soggetto all'Amministrazione Regionale. Più Regioni compongono lo Stato.	Definizione Ambito territoriale soggetto all'Amministrazione Provinciale. Più Province compongono una Regione.
Attributi	Attributi
Attributi della classe	Attributi della classe
09010601 REGION_IST codice istat regione NumericString(16) Codice Istat Regione	09010501 PROVIN_IST codice istat provincia NumericString(16) Codice Istat Provincia
09010602 REGION_NOM nome regione [1..*] Multilinguismo (DataType) Nome Regione	09010502 PROVIN_NOM nome provincia [1..*] Multilinguismo (DataType) Nome Provincia
Componenti spaziali della classe	Componenti spaziali della classe
090106101 REGION_EXT Estensione GU_CSsurface2D - Complex Surface 2D Si ottiene come aggregazione del territorio delle provincie che la compongono	090105101 PROVIN_EXT Estensione GU_CSsurface2D - Complex Surface 2D Si ottiene come aggregazione del territorio di comuni che la compongono
090106102 REGION_SED Sede amministrativa GU_Point2D - Point 2D Individua la Sede Amministrativa; deve ricadere all'interno dell'edificio comprendente la Sede Amministrativa dell'Ente	090105102 PROVIN_SED Sede amministrativa GU_Point2D - Point 2D Individua la Sede Amministrativa; deve ricadere all'interno dell'edificio comprendente la Sede Amministrativa dell'Ente
Ruoli	Ruoli
PvIrreg Definisce le provincie di cui è composta la specifica regione. Deve essere verificato se qualunque regione italiana è composta da almeno una provincia o se viceversa situazioni particolari in cui all'interno della regione non esiste neppure una provincia PvIrreg [1..*] : PROVIN <i>inverso</i> RgIprov [1]	RgIprov Definisce la regione cui appartiene la specifica provincia RgIprov [1] : REGION <i>inverso</i> PvIrreg [1..*] ComIprov Definisce di quali comuni è composta una specifica provincia ComIprov [1..*] : COMUNE <i>inverso</i> Pvdicm [1]

Figura 3. Classe Regione e Provincia del DBT.

Da questo esempio si può capire che per molte classi esiste una differenza sostanziale di struttura dati tra INSPIRE e DBT.

3. La conversione dati

Come accennato in precedenza, in molti sistemi la base dati è integrata con applicativi che possono essere sostituiti solo a seguito di una lunga analisi e sviluppo di nuovo codice. Questo implica che, in molti casi, non è immediato convertire la base dati esistente ma è necessario utilizzare un approccio diverso.

Sulla base dell'esperienza fatta, Esri ha definito un workflow per la conversione dei dati verso il modello INSPIRE. Questo workflow si basa su una serie di strumenti che rispondono pienamente alle necessità di INSPIRE e coprono l'intero disegno architetturale, come mostrato in Figura 4

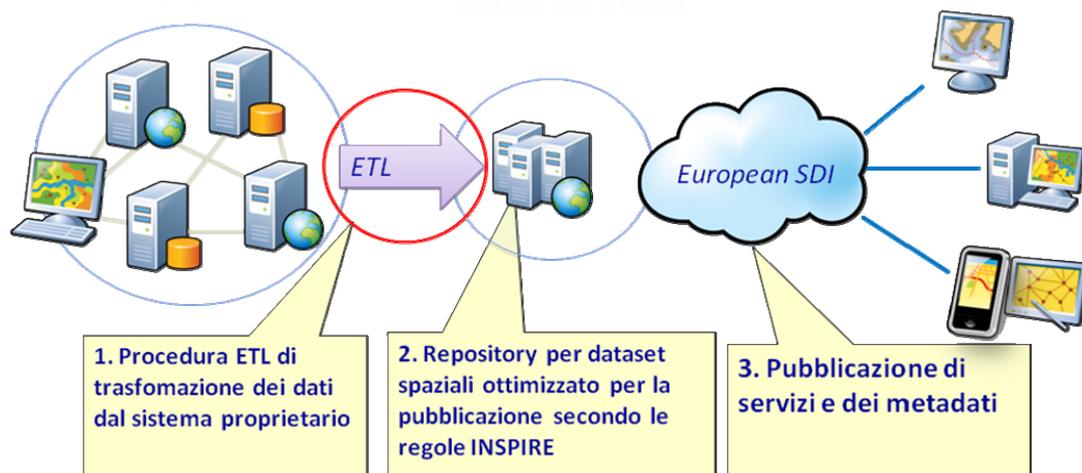


Figura 4. Workflow di trasformazione dati.

In questo workflow l'unico elemento che deve essere implementato è cerchiato in rosso, e cioè la conversione dati dal sistema proprietario verso il geodatabase definito da *ArcGIS for INSPIRE*, il prodotto Esri per la gestione dati e pubblicazione dei servizi secondo le regole INSPIRE.

Questa trasformazione però permette di ottemperare agli obblighi INSPIRE ma non a quelli del DBT. E' quindi necessario creare un'altra trasformazione, e quindi la situazione è quella mostrata in *Figura 5*:

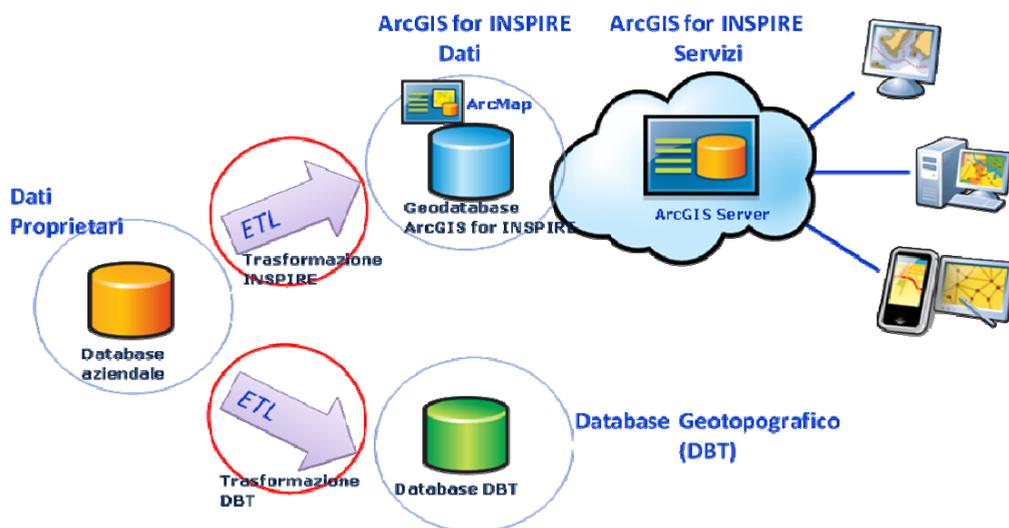


Figura 5. Schema delle trasformazioni dati.

Cioè è necessario creare due trasformazioni complesse e diverse per convertire i dati aziendali nei due formati richiesti, quello di INSPIRE e quello del DBT. Nel prossimo paragrafo vedremo come sia possibile limitare la complessità ad una sola trasformazione.

4. La conversione DBT-INSPIRE

In generale una conversione dati può essere definita univocamente quando sono noti gli schemi di partenza e di destinazione. In effetti questo è il caso di DBT ed INSPIRE: a parte particolarità semantiche, che possono comunque essere risolte in una fase di analisi, è possibile creare una trasformazione DBT→INSPIRE almeno per i temi di interesse per entrambi gli standard.

Si costruisce quindi una trasformazione come illustrato in *Figura 6*:

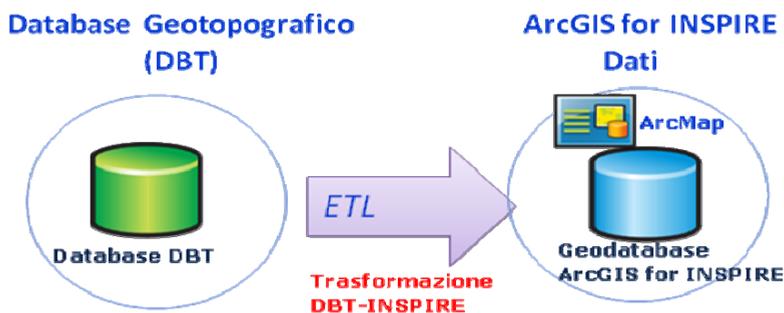


Figura 6. Trasformazione DBT==>INSPIRE.

La scelta di trasformare da DBT ad INSPIRE e non viceversa è dovuta sia al maggior dettaglio costituito da DBT ma soprattutto perché il Decreto pubblicato in Gazzetta impone che le nuove acquisizioni siano comunque nel formato DBT, mentre INSPIRE impone solo la disponibilità di dati e servizi nel formato indicato. Si ha quindi un obbligo di esistenza più forte nei riguardi del DBT.

Questa trasformazione ci permette di ridisegnare lo schema di *Figura 5* nel seguente modo:

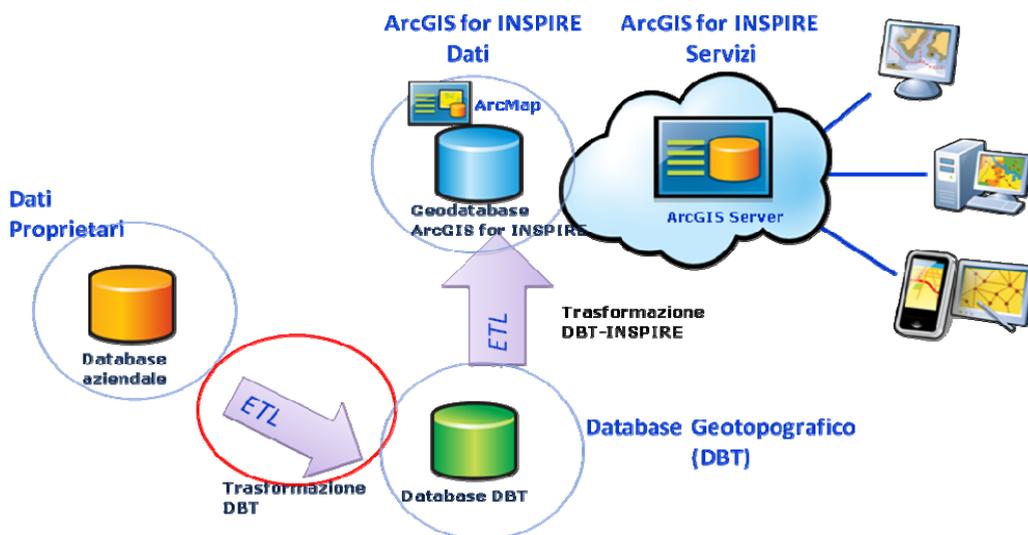


Figura 7. Trasformazione dati semplificata.

Dove si vede che la trasformazione cerchiata in rosso è l'unico elemento che deve essere implementato, mentre la trasformazione DBT-INSPIRE rimane invariante per ogni ambiente.

La realizzazione di questo schema è possibile con le seguenti condizioni:

- lo schema DBT deve essere per la maggior parte un soprainsieme di quello INSPIRE
- la trasformazione deve essere fattibile
- deve essere disponibile uno strumento ETL che permetta la trasformazione

Il primo punto può essere dedotto sia dalla considerazione fatta in precedenza che il DBT si occupa più in dettaglio della realtà italiana (come mostrato per quanto riguarda le unità amministrative), sia da un'analisi sui temi comuni, quali ad esempio l'Idrografia. In altre parole si può dire che, avendo il DBT come obiettivo quello di costituire la base di riferimento per i dati presenti in Italia, comunque non sarebbe possibile avere ulteriori informazioni se pure richieste da INSPIRE.

Il secondo punto riguarda la *fattibilità* della trasformazione, cioè la possibilità di definire una conversione dall'origine alla destinazione che possa essere eseguita in tempi computazionalmente accettabili. Dall'analisi effettuata risulta che per tutte le trasformazioni, per quanto complesse, esiste un algoritmo di trasformazione. Esempi di trasformazione sono: conversione di attributi, split di valori in più campi, trasformazione di geometria, creazione di più geometrie semplici a partire da una complessa, etc.

Infine con *disponibilità* intendiamo la possibilità di avere un prodotto di ETL con costi che siano adeguati ai vari progetti in essere. Per quest'ultimo punto è possibile esaminare il prodotto FME (Feature Manipulation Engine) della Safe Software, prodotto incluso nell'estensione Data Interoperability di ArcGIS Desktop di Esri.

Questo prodotto ha già con sé più di cento trasformatori, cioè operatori logici, alfanumerici, geometrici che, composti attraverso un legame configurabile, costituiscono un workflow di trasformazione, come evidenziato in *Figura 8*

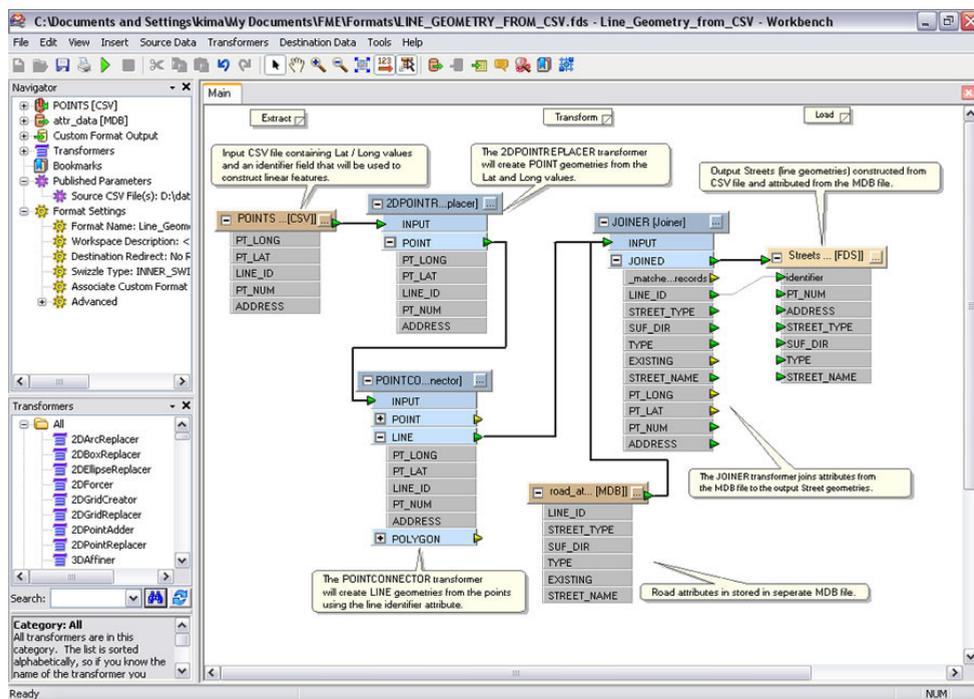


Figura 8. Esempio di Workbench FME.

Dato che è possibile salvare e rilanciare il modello di trasformazione, anche in modalità batch, è possibile mantenere allineate le basi dati senza necessità di ulteriori interventi manuali.

Utilizzando questo strumento Esri Italia sta definendo le trasformazioni necessarie da DBT ad INSPIRE: è già stata implementata la trasformazione per gli Ambiti Amministrativi e per l'Idrografia, e a breve saranno implementate altre Classi.

5. Conclusioni

La fase di implementazione delle Data Specifications di INSPIRE si sta sovrapponendo con l'implementazione del Database Geotopografico definito dalle specifiche Italiane pubblicate in Gazzetta Ufficiale a Febbraio 2012. Questo scenario presuppone per i soggetti interessati una doppia trasformazione di dati, con conseguenti costi da affrontare in termini di formazione e sviluppi software. La soluzione proposta da Esri Italia prevede una trasformazione DBT → INSPIRE che, unita ai prodotti Esri rivolti ad INSPIRE, permette di ridurre notevolmente l'impegno necessario per un'azienda per convertire i propri dati nei formati richiesti. La trasformazione, in continuo sviluppo man mano che i temi INSPIRE saranno consolidati, può essere anche modificata dagli utilizzatori in modo da fornire una soluzione il più possibile aderente sia agli standard che alle esigenze di ogni sistema.