

INSPIRE COME OPPORTUNITÀ: ARMONIZZAZIONE DEI DATASET PER L'INTEROPERABILITÀ A LIVELLO REGIONALE

DATI ISTAT E TRASFORMAZIONI COMPLESSE

Webinar – 10 dicembre 2013

Jody Marca – jmarca@gmail.com



OBIETTIVI

- Definire e testare un set di trasformazioni in ambiente Open Source
- Generare un set di dati GML conforme agli XSD di INSPIRE il cui contenuto sia il più completo possibile
- Portare alla luce le problematiche relative alla produzione e visualizzazione di dati GML complessi e di grandi dimensioni

DATI UTILIZZATI

Sono stati scaricati gli shapefile relativi ai confini amministrativi con validità 1° gennaio 2011

Gli shapefile contengono solo gli elementi poligonali che rappresentano i comuni, le province e le regioni italiane.

Sono assenti i limiti lineari, la geometria poligonale che rappresenta la Repubblica Italiana e le geometrie puntiformi che rappresentano le sedi amministrative

Download dati: <http://www.istat.it/it/archivio/24580>

STEP 1: ELABORAZIONE DEI DATI SCARICATI

La prima fase della trasformazione è stata quella relativa alla generazione dei dati di input mancanti

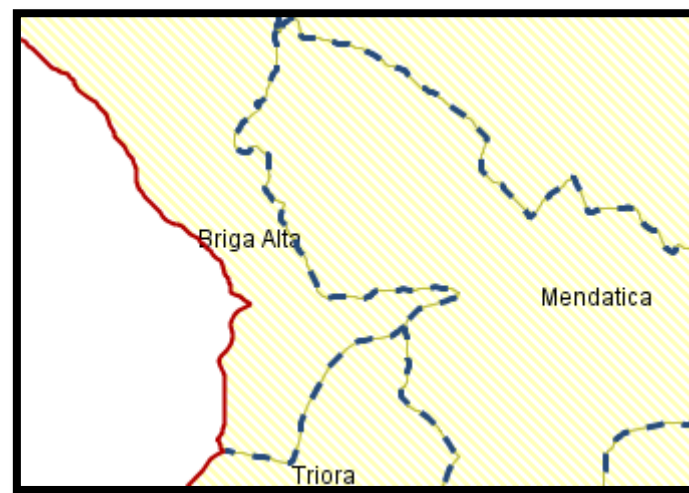
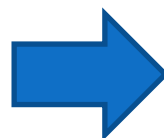
Si è generato il poligono 2D rappresentante la Repubblica Italiana operando una unione / dissolve di tutte le geometrie poligonali delle regioni italiane

Si sono generate le sedi puntiformi utilizzando una la funzione point on surface che, a differenza del calcolo del centroide, restituisce sempre un punto interno alla superficie del poligono.

Queste sedi sono ovviamente virtuali ma sono state introdotte per dare completezza all'esempio

STEP 1: ELABORAZIONE DEI DATI SCARICATI /2

Si sono infine generati i limiti lineari che rappresentano i boundary del comune, ovviamente alcuni di questi sono condivisi tra più geometrie poligonali mentre altri sono esclusivi (come nel caso dei limiti statali e delle acque territoriali)



Le linee blu tratteggiate rappresentano i limiti condivisi mentre quelle rosse continue i limiti esclusivi

STEP 2: CARICAMENTO DEI DATI

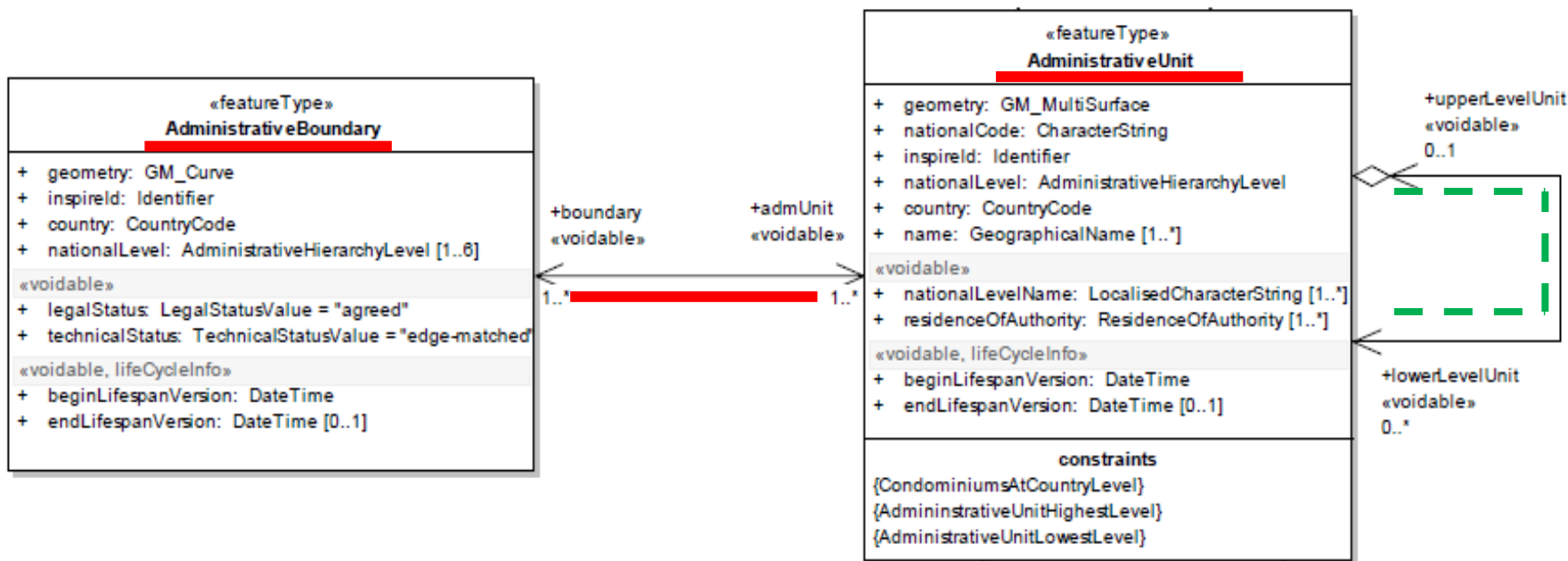
La seconda fase è stata quella di caricamento degli shape sia scaricati sia derivati in un database PostgreSQL con estensione spaziale PostGIS

Si sono caricati i dati attraverso il tool shp2pgsql che crea una tabella per ogni shape

Il sistema di riferimento utilizzato in caricamento è stato ED50 / UTM zone 32N (EPSG:23032) che era quello associato agli shapefile scaricati; le geometrie sono poi state riproiettate utilizzando la funzione di PostGIS (basate sulle librerie Proj4) nel sistema di riferimento ETRS89 / UTM zone 32N (EPSG:25832)

STEP 3: MAPPING DEI DATI

La terza fase è stata quella di definire lo schema delle tabelle utili a rappresentare le feature INSPIRE



Tre tabelle per Administrative Unit, Administrative Boundary e la tabella n-n per legare le due entità
L'autorelazione su Administrative Unit, essendo 0..n, è stata resa con un semplice attributo

STEP 3: ADMINISTRATIVE UNIT

Popolata con tutti gli i confini poligonali e le sedi puntiformi di stato, regioni, provincie e comuni. Durante il popolamento si è inoltre compilato (in base al tipo di unità amministrativa) il valore del ordine

Il campo idpadre è usato per esplicitare l'auto relazione

Sono presenti due attributi di tipo nome utili ad implementare il multilinguismo poiché nei dati i comuni della provincia autonoma di Bolzano erano riportati sia in italiano che in tedesco.

```
administrativeunit
gmlid varchar(20) NOT NULL (PK)
localid varchar(10) NOT NULL
idpadre varchar(20) NULL
ordine int4(10) NOT NULL
nomeordine varchar(100) NOT NULL
nome_ita varchar(100) NOT NULL
nome_ted varchar(100) NULL
geom geometry(2147483647) NULL
sede geometry(2147483647) NULL
```

Mapping valori attributo ordine

Repubblica	1° ordine
Regione	2° ordine
Provincia	3° ordine
Comune	4° ordine

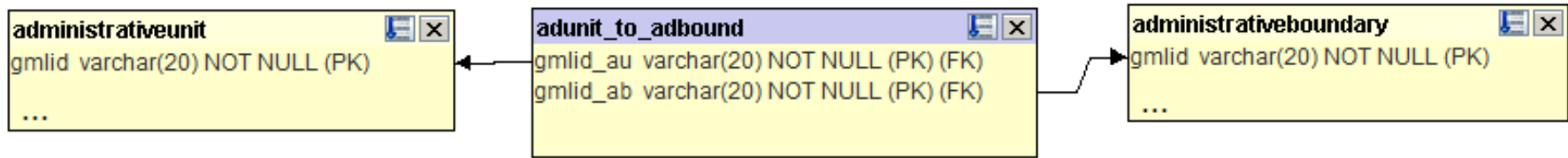
STEP 3: ADMINISTRATIVE BOUNDARY

Contiene tutti gli elementi lineari generati partendo dai poligoni che rappresentano il territorio delle unità amministrative.

```
administrativeboundary
gmlid varchar(20) NOT NULL (PK)
localid varchar(10) NOT NULL
ordine int4(10) NOT NULL
geom geometry(2147483647) NULL
```

Durante il popolamento, come per la tabella `administrativeunit`, si è compilato il valore di ordine. Si ricorda che la geometria riportata nella tabella è una `Linestring` semplice quindi nel caso di due comuni con un limite condiviso rappresentato da una `Multilinestring` in questa tabella sarà necessario inserire un record per ogni componente della `Multilinestring`.

STEP 3: RELAZIONE TRA UNIT E BOUNDARY



La tabella `adunit_to_adbound` permette di associare ogni geometria poligonale alle relative geometrie lineari che ne rappresentano il boundary e vice versa.

La compilazione di questa tabella è stata fatta con una query spaziale nella quale si richiedeva che la geometrie lineari fossero completamente contenute nel boundary della geometria poligonale e che il valore dell'attributo `ordine` per la tabella `administrativeunit` fosse uguale a quello della tabella `administrativeboundary`

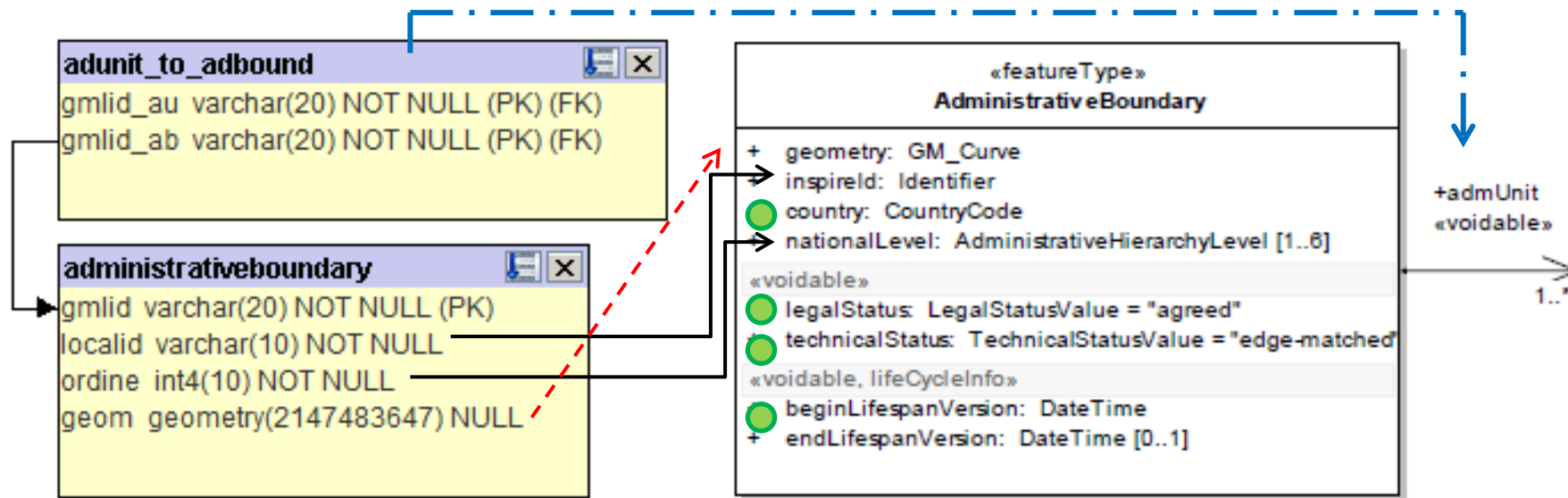
STEP 4: ESTRATTORE GML

La fase di estrazione poteva essere condotta attraverso un servizio WFS, utilizzando un ETL (GeoKettle o FME) o con codice scritto ad-hoc per produrre un file GML. Si è scelta questa ultima strada poiché aveva alcuni pregi:

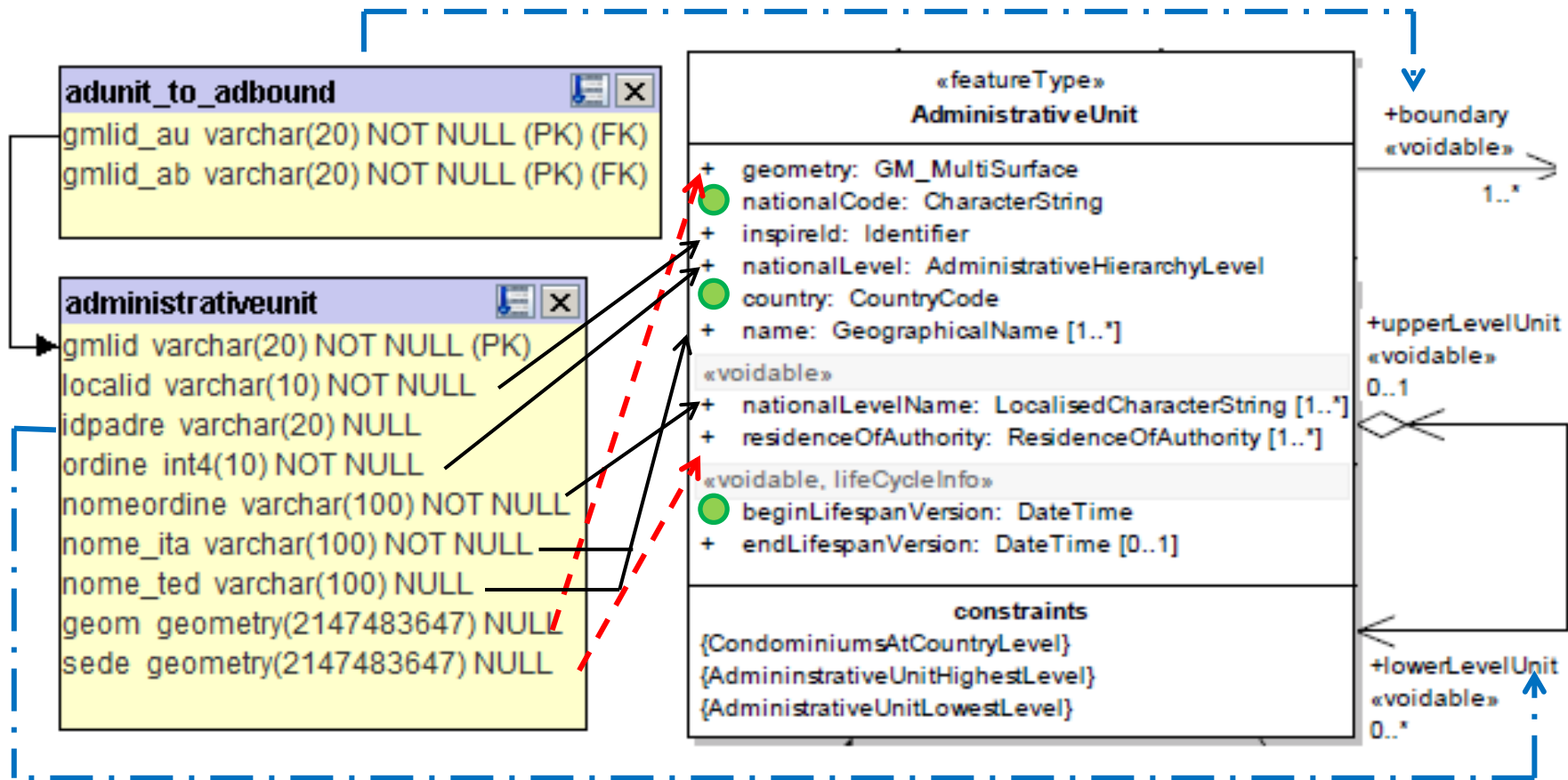
- permette la produzione di grandi file GML (centinaia di MB) senza problemi di memoria
- molto flessibile ai cambiamenti per produrre prove confrontabili
- altamente configurabile poiché essendo il mapping definito dall'estrattore è stato possibile gestire quali associazioni gestire by-value e quali by-reference

STEP 4: ESTRATTORE GML /2

Il compito dell'estrattore è stato quello di estrarre il dato dal database, strutturarlo in un GML compliant al XSD di INSPIRE e contestualmente aggiungere i valori ritenuti omogenei a tutti i dati (di seguito etichettati con un pallino verde).



STEP 4: ESTRATTORE GML /3



Le frecce nere esplicitano il mapping tra attributi, quelle rosse tra geometrie e quelle azzurre le relazioni (implementate attraverso degli X-link)

STEP 4: ESTRATTORE GML /4

Durante lo sviluppo dell'estrattore si è deciso di rendere possibile configurabile:

- il numero massimo di decimali da utilizzare durante la scrittura delle coordinate GML
- se estrarre tutti i livelli (comuni, province, regioni e stato) o solo uno di essi
- se, durante l'estrazione delle Administrative Unit, estrarre anche le geometrie puntiformi delle sedi o solo le geometrie poligonali dell'estensione comunale

STEP 4: ESTRATTORE GML /5

Al termine del processo di estrazione si è validato il GML prodotto ma, a causa delle grandi dimensioni, non è stato possibile utilizzare i più comuni software di editing XML

Validazione rispetto allo schema:

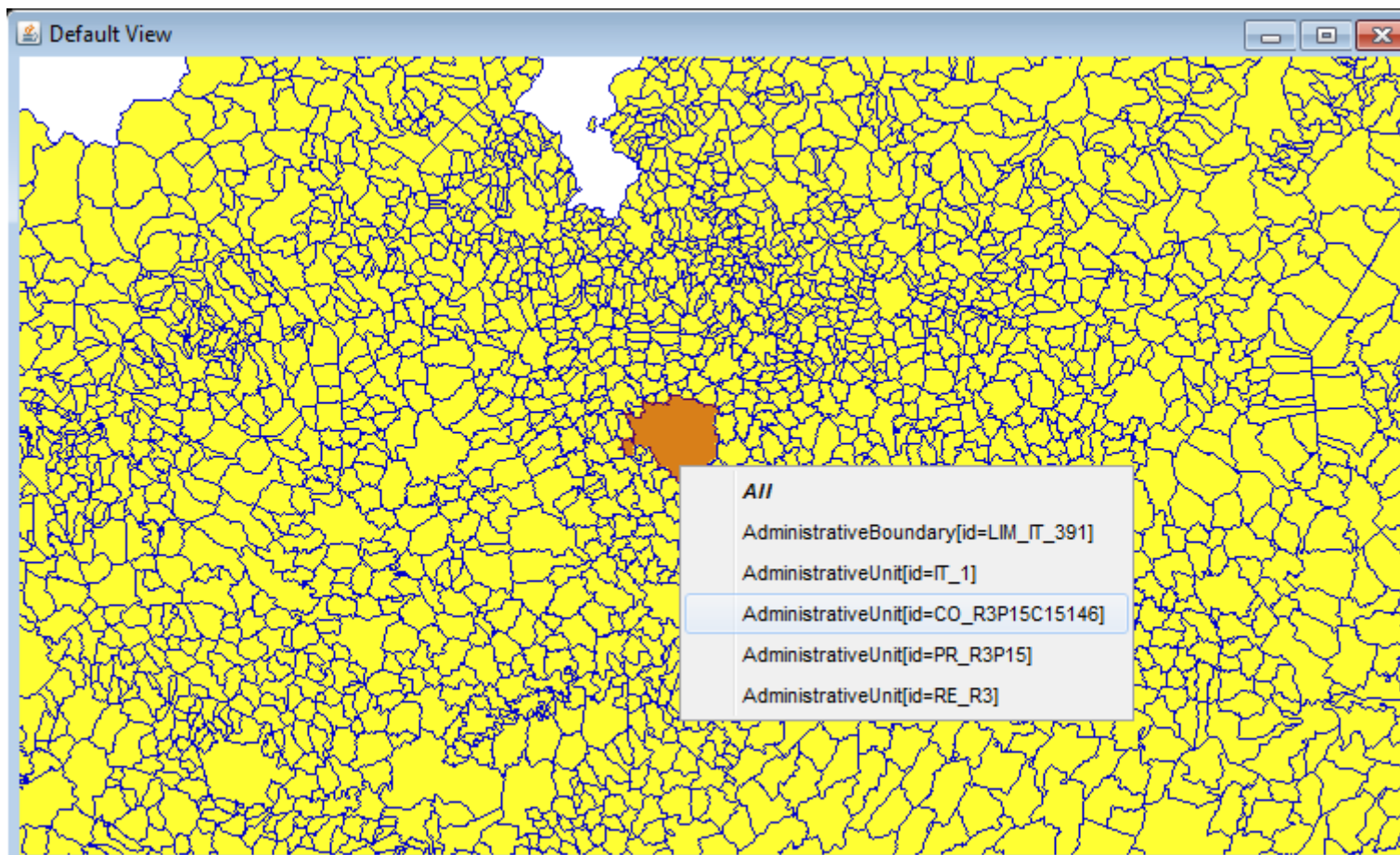
<http://inspire.ec.europa.eu/schemas/au/3.0/AdministrativeUnits.xsd>

```
Italia_3dec.gml is valid
Italia_5dec.gml is valid
Italia_8dec.gml is valid
Italia_con_sedi_3dec.gml is valid
Italia_con_sedi_5dec.gml is valid
Italia_con_sedi_8dec.gml is valid
Solo_Comuni_3dec.gml is valid
Solo_Comuni_5dec.gml is valid
Solo_Comuni_8dec.gml is valid
Solo_Province_3dec.gml is valid
Solo_Province_5dec.gml is valid
Solo_Province_8dec.gml is valid
Solo_Regioni_3dec.gml is valid
Solo_Regioni_5dec.gml is valid
Solo_Regioni_8dec.gml is valid
Solo_Repubblica_3dec.gml is valid
Solo_Repubblica_5dec.gml is valid
Solo_Repubblica_8dec.gml is valid
Validazione terminata in: 887 secondi
```

Si è implementata una classe JAVA che esegue tale validazione con la libreria Apache Xalan

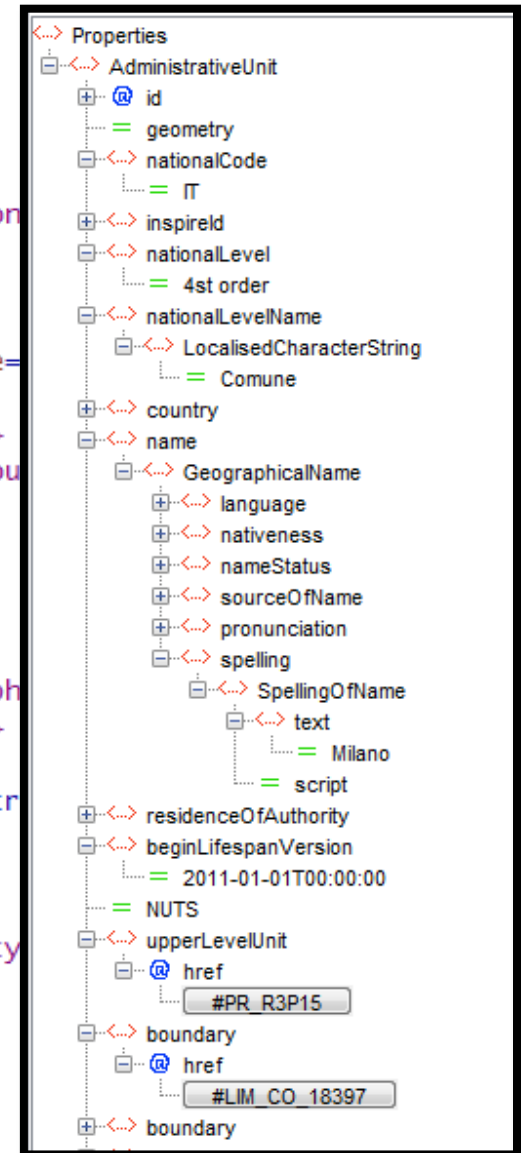
STEP 5: VISUALIZZAZIONE GML

Per la visualizzazione del GML estratto si è utilizzato il GML Viewer della Snowflake Software poiché permette una piena navigazione degli Xlink



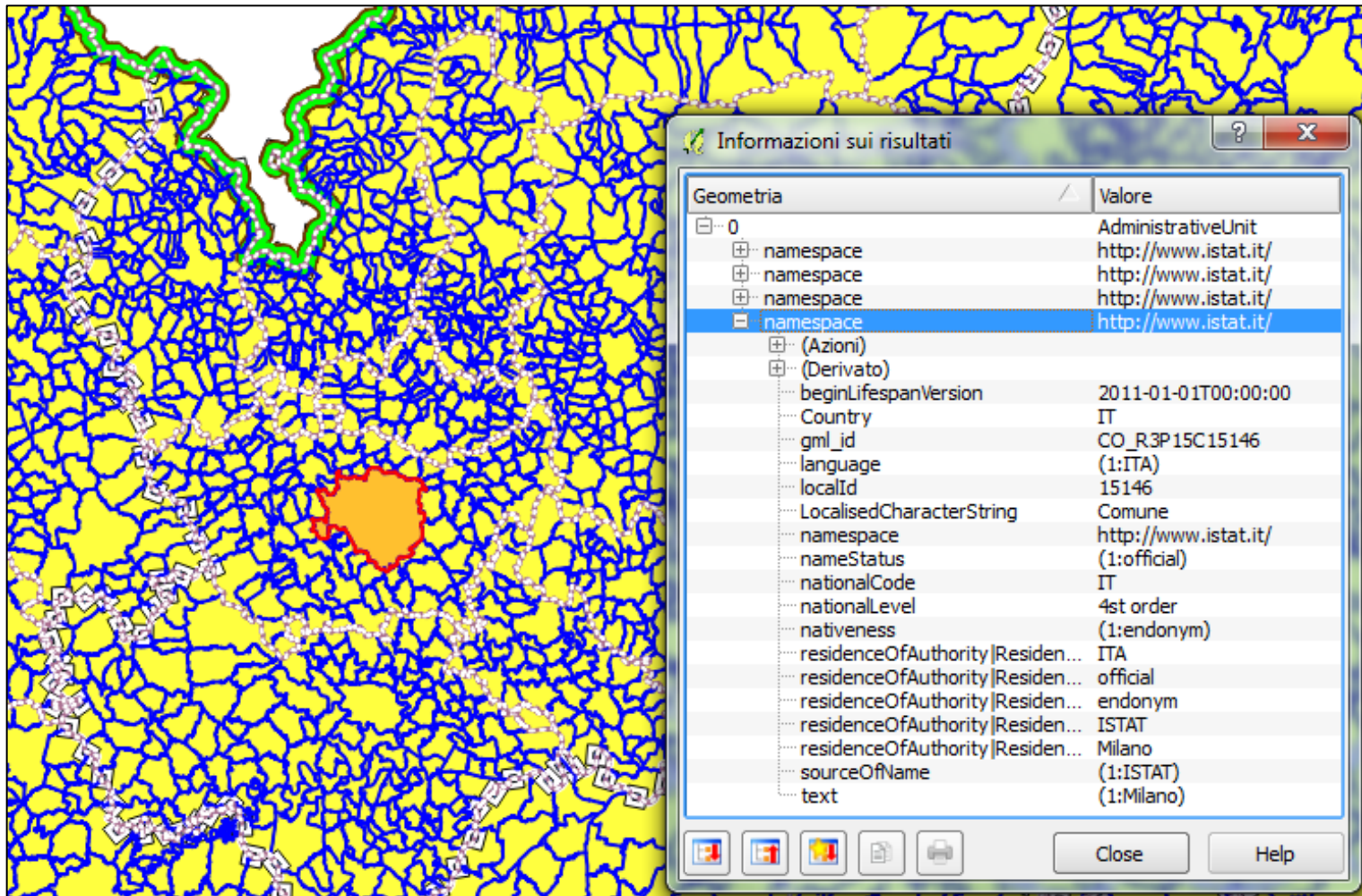
STEP 5: VISUALIZZAZIONE GML / 2

```
<au:AdministrativeUnit gml:id="CO_R3P15C15146">
  <au:geometry>OMESSA</au:geometry>
  <au:nationalCode>IT</au:nationalCode>
  <au:inspireId><base:Identifier><base:localId>15146</base:localId>
    <base:namespace>http://www.istat.it/</base:namespace>
  </base:Identifier></au:inspireId><au:nationalLevel>4st order</au:nation
  <au:nationalLevelName>
    <gmd:LocalisedCharacterString>Comune</gmd:LocalisedCharacterString>
  </au:nationalLevelName><au:country>
    <gmd:Country codeList="...ML_gmxCodeLists.xml#Country" codeListValue=
  </au:country><au:name><gn:GeographicalName>
    <gn:language>ITA</gn:language><gn:nativeness>endonym</gn:nativeness>
    <gn:nameStatus>official</gn:nameStatus><gn:sourceOfName>ISTAT</gn:sou
    <gn:pronunciation xsi:nil="true"/>
    <gn:spelling><gn:SpellingOfName>
      <gn:text>Milano</gn:text><gn:script/>
    </gn:SpellingOfName></gn:spelling>
  </gn:GeographicalName></au:name>
  <au:residenceOfAuthority><au:ResidenceOfAuthority><au:name><gn:Geograph
    <gn:language>ITA</gn:language><gn:nativeness>endonym</gn:nativeness>
    <gn:nameStatus>official</gn:nameStatus>
    <gn:sourceOfName>ISTAT</gn:sourceOfName><gn:pronunciation xsi:nil="tr
    <gn:spelling><gn:SpellingOfName>
      <gn:text>Milano</gn:text><gn:script/>
    </gn:SpellingOfName></gn:spelling>
  </gn:GeographicalName></au:name><au:geometry/></au:ResidenceOfAuthority>
</au:residenceOfAuthority>
<au:beginLifespanVersion>2011-01-01T00:00:00</au:beginLifespanVersion>
<au:NUTS/>
<au:boundary xlink:href="#LIM_CO_18397"/>
<au:boundary xlink:href="#LIM_CO_18700"/>
```



STEP 5: VISUALIZZAZIONE GML / 3

Anche QGIS 2 che permette la visualizzazione del GML (anche se con alcune limitazioni)



STEP 5: VISUALIZZAZIONE GML / 4

QGIS 2, a differenza di GML Viewer, permette una maggior configurazione nella presentazione delle geometrie caricate, inoltre permette di elaborare i dati caricati in GML e salvarli in altri formati; tuttavia sono state riscontrate le seguenti limitazioni:

1. Non permette la navigazione degli X-Link tra le feature GML (sono infatti rimosse tutte le associazioni presenti)
2. Supporto embrionale degli attributi multi valore (concatena i valori in un insieme di stringhe)
3. Supporto di una sola geometria per feature GML (quindi se sono presenti le sedi puntiformi non vengono visualizzate come geometrie ma sono tradotte in WKT e presentate come attributi)

FLUSSO DI TRASFORMAZIONE

Di seguito è riportata una figura riassuntiva di tutto il processo che ha permesso di trasformare i dati ISTAT dal formato shapefile in files GML conformi al XSD degli Administrative units definiti da INSPIRE

Elaborazioni
su shape ISTAT



Caricamento
In PostGIS

Mapping ed
elaborazioni
in PostGIS





















Estrattore
GML in Java

Visualizzazione
GML prodotto



CONSIDERAZIONI SUL GML

La possibilità di definire a quale coordinata decimale “troncare” le coordinate in GML ha permesso di confrontare le dimensioni dei file prodotti.

Nome	Dimensione	Nome	Dimensione	Nome	Dimensione
 Italia_3dec.gml	307.055 KB	 Solo_Comuni_3dec.gml	227.553 KB	 Solo_Regioni_3dec.gml	23.754 KB
 Italia_5dec.gml	348.949 KB	 Solo_Comuni_5dec.gml	256.753 KB	 Solo_Regioni_5dec.gml	27.594 KB
 Italia_8dec.gml	411.748 KB	 Solo_Comuni_8dec.gml	300.521 KB	 Solo_Regioni_8dec.gml	33.352 KB
 Italia_con_sedi_3dec.gml	308.677 KB	 Solo_Province_3dec.gml	37.186 KB	 Solo_Repubblica_3dec.gml	17.751 KB
 Italia_con_sedi_5dec.gml	350.603 KB	 Solo_Province_5dec.gml	43.185 KB	 Solo_Repubblica_5dec.gml	20.607 KB
 Italia_con_sedi_8dec.gml	413.450 KB	 Solo_Province_8dec.gml	52.174 KB	 Solo_Repubblica_8dec.gml	24.890 KB

In Italia_con_sedi.gml sono contenuti tutti i dati mentre Italia sono contenute solo le geometrie poligonali e lineari; gli altri files contengono solo un sottoinsieme dei dati. Il suffisso *_dec indica il numero di decimali a cui sono troncate le coordinate (il default in PostGIS è settato a 15).

CONSIDERAZIONI SUL GML / 2

File di queste dimensioni sono difficilmente gestibili dai visualizzatori ed inoltre possono essere anche problematici da produrre.

È possibile ridurre sensibilmente la dimensione dei files utilizzando pesantemente gli X-link, non solo per implementare le relazioni, ma anche all'interno della definizione della geometria. Si potrebbe definire le geometrie delle unità amministrative richiamando le componenti lineari dei boundary, e le componenti a livello di province, regioni e stato richiamando i boundary a livello comunale .

Sfortunatamente questo approccio sembra non essere ancora supportato dai principali visualizzatori di GML 3.2.1

DIFFICOLTÀ DI PRODUZIONE DEL GML

È stato necessario aggiornare Postgis all'ultima versione 2.1 poiché le versioni precedenti non producevano le geometrie con l'attributo gml:id diventato obbligatorio con la versione 3.2.1 del GML

Sfortunatamente nel caso dei poligoni ci sono ancora dei problemi di duplicazione che hanno richiesto un post processing durante la fase di estrazione ma si presume che questa problematica possa essere risolta nelle future versioni.

```
<gml:MultiSurface srsName="urn:ogc:def:crs:EPSG::3035" gml:id="id8">  
  <gml:surfaceMember><gml:Polygon gml:id="id8">  
    ...  
  </gml:Polygon></gml:surfaceMember>  
</gml:MultiSurface>
```