

Guida all'uso del web-service SDMX

Introduzione

L'obiettivo di questo documento è l'illustrazione sintetica degli step che tecnicamente bisogna compiere affinché un generico client sia in grado di interagire con il web service SDMX dell'Istat.

Nel documento verrà fatto riferimento a nozioni SDMX sinteticamente descritte nell'appendice per eventuali approfondimenti si rimanda alla lettura della documentazione SDMX disponibile sul sito <http://sdmx.org>.

Step funzionali

Gli step da effettuare per utilizzare il web service sono:

1. Effettuare la connessione al web service SDMX (tramite l'utilizzo del relativo end-point che è presente nella pagina di registrazione) richiamando il metodo "QueryStructure" del web service usando la query contenuta nel file "get_all_dataflow.xml". Tale metodo permette di ottenere tutti i dataflow (dataset) disponibili. Il web-service restituirà una stringa xml contenente l'elenco di tutti i dataflow (dataset).

Le informazioni da reperire sono le seguenti:

- Gli attributi *id*, *version*, *agencyID* del tag¹ <structure:dataflow>, che identificano univocamente i dataflow
- Il contenuto dei tag <structure:Name xml:lang="en"> ed <structure:Name xml:lang="it">, figli del tag <structure:dataflow>: che corrispondono ai nomi italiano ed inglese dei dataflow
- Per ogni tag <structure:Dataflow>, il contenuto del tag <structure:KeyFamilyRef> rappresentato dai tag <structure:KeyFamilyID>, <structure:KeyFamilyAgencyID>, <structure:KeyFamilyVersion>: che identificano univocamente la keyfamily o data structure definition (DSD) relativa al dataflow . La DSD rappresenta la struttura del dataset in termini di dimensioni e relative codelist, attributi e misure.

```
<structure:Dataflow id="144_35" version="1.0" agencyID="IT1" isFinal="true">
  <structure:Name xml:lang="en">Consumer price index for the whole nation - weights (NIC - from 2011 onwards)</structure:Name>
  <structure:Name xml:lang="it">Indice dei prezzi al consumo per l intera collettivita' - pesi (NIC - dal 2011) </structure:Name>
  <structure:KeyFamilyRef>
    <structure:KeyFamilyID>SEP_PREZZI_NIC_FOI_WEIGHT_2011</structure:KeyFamilyID>
    <structure:KeyFamilyAgencyID>IT1</structure:KeyFamilyAgencyID>
    <structure:Version>1.0</structure:Version>
  </structure:KeyFamilyRef>
```

Fig.1: Dettaglio relativo al dataflow nell'SDMX-Response alla query "get_all_dataflow.xml"

¹ Un **tag** è una parola chiave o un termine associato a un'informazione, che descrive l'oggetto rendendo possibile la classificazione e la ricerca di informazioni basata su parole chiave. I tag sono generalmente scelti in base a criteri informali e personalmente dagli autori/creatori dell'oggetto dell'indicizzazione. Tipicamente a ogni oggetto sono applicati più tag (da wikipedia)

2. Estrazione della Data Structure Definition (metadati del Dataflow/Dataset)

Per conoscere la struttura del dataflow si deve richiamare sempre il metodo "QueryStructure" del web-service usando in questo caso la query "get_a_keyfamily.xml" dopo aver effettuato le seguenti sostituzioni:

- Nel tag <registry:KeyFamilyID> deve essere inserito l'id della keyfamily
- Nel tag <registry:AgencyID> deve essere indicato il codice dell'organizzazione (nel caso dell'ISTAT è di norma "IT1")
- Nel tag <registry:Version> deve essere indicata la versione.

```
<registry:KeyFamilyRef>
  <registry:AgencyID>IT1</registry:AgencyID>
  <registry:KeyFamilyID>SEP_IND_COSTR_F0</registry:KeyFamilyID>
  <registry:Version>1.2</registry:Version>
</registry:KeyFamilyRef>
```

Fig.2: Dettaglio relativo alla query "SDMX_Compact_Data_Query_Example.xml" con l'indicazione del tag in cui inserire l'id del dataflow di interesse

Il risultato della query restituisce è rappresentato dai metadati strutturali associati al dataflow di interesse, ed in particolare:

- Le dimensioni (tag <structure:Dimension> e <structure:TimeDimension>), gli attributi (tag <structure:Attribute>) e le misure (tag <structure:PrimaryMeasure>) associati al dataflow/dataset considerato: tali tag sono figli del tag <structure:Components>, a sua volta figlio del tag <structure:KeyFamily>

```
<structure:Components>
<structure:Dimension conceptRef="FREQ" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Dimension conceptRef="REF_AREA" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Dimension conceptRef="ADJUSTMENT" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Dimension conceptRef="INDICATOR" conceptSchemeRef="SEP_INDCOSTR_COMPACT" conceptSchemeAgency="IT1" conceptSchemeVersion="1.
<structure:Dimension conceptRef="ACTIVITY" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Dimension conceptRef="BASE_PER" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Dimension conceptRef="MARKET" conceptSchemeRef="SEP_INDCOSTR_COMPACT" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0"
<structure:Dimension conceptRef="UN_MEAS" conceptSchemeRef="SEP_INDCOSTR_COMPACT" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0"
<structure:TimeDimension conceptRef="TIME_PERIOD" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0"
<structure:PrimaryMeasure conceptRef="OBS_VALUE" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0"
<structure:Attribute conceptRef="COMMENT" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" attachme
<structure:Attribute conceptRef="CONF_STATUS" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" cod
<structure:Attribute conceptRef="OBS_PRE_BREAK" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" a
<structure:Attribute conceptRef="OBS_STATUS" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" code
<structure:Attribute conceptRef="DECIMALS" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" codeList="
<structure:Attribute conceptRef="BREAK" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" attachme
<structure:Attribute conceptRef="COLLECTION" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" code
<structure:Attribute conceptRef="TIME_FORMAT" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" cod
<structure:Attribute conceptRef="TITLE" conceptSchemeRef="CROSS_DOMAIN" conceptSchemeAgency="IT1" conceptSchemeVersion="1.0" attachme
</structure:Components>
```

Fig.3: dettaglio della stringa della DSD contenente la keyfamily: ovvero l'elenco di dimensioni, attributi, misure

- Le codelist (tag `<structure:CodeList>` figli del tag `<registry:CodeLists>`) associate alle dimensioni e/o agli attributi, ed i relativi codici (tag `<structure:Code>`)

```

<registry:CodeLists>
  <structure:CodeList id="CL_TIME_FORMAT" version="1.0" agencyID="IT1" isFinal="true">
    <structure:Name xml:lang="it">Formato tempo</structure:Name>
    <structure:Name xml:lang="en">Time Format</structure:Name>
    <structure:Code value="PIY">
      <structure:Description xml:lang="it">Annual</structure:Description>
      <structure:Description xml:lang="en">Annual</structure:Description>
    </structure:Code>
    <structure:Code value="PGM">
      <structure:Description xml:lang="it">Semi-Annual</structure:Description>
      <structure:Description xml:lang="en">Semi-Annual</structure:Description>
    </structure:Code>
    <structure:Code value="P3M">
      <structure:Description xml:lang="it">Quarterly</structure:Description>
      <structure:Description xml:lang="en">Quarterly</structure:Description>
    </structure:Code>
    <structure:Code value="P1M">
      <structure:Description xml:lang="it">Monthly</structure:Description>
      <structure:Description xml:lang="en">Monthly</structure:Description>
    </structure:Code>
  </structure:CodeList>

```

Fig.4: dettaglio della stringa della DSD contenente un esempio di codelist con i relativi codici

- I ConceptScheme (o elenchi di concetti, tag `<structure:ConceptScheme>` figli del tag `<registry:Concepts>`) i cui concetti (tag `<structure:Concept>`) sono associati alle dimensioni e agli attributi

```

<registry:Concepts>
  <structure:ConceptScheme id="SEP_INDCOSTR_COMPACT" version="1.0" agencyID="IT1" isFinal="true">
    <structure:Name xml:lang="it">Industria e costruzioni</structure:Name>
    <structure:Name xml:lang="en">Industry and construction</structure:Name>
    <structure:Concept id="INDICATOR">
      <structure:Name xml:lang="en">STS Indicator</structure:Name>
      <structure:Description xml:lang="en">STS Indicator</structure:Description>
    </structure:Concept>
    <structure:Concept id="MARKET">
      <structure:Name xml:lang="it">Mercato di riferimento</structure:Name>
      <structure:Name xml:lang="en">Market</structure:Name>
    </structure:Concept>
    <structure:Concept id="UH_MEAS">
      <structure:Name xml:lang="it">Misura</structure:Name>
      <structure:Name xml:lang="en">Measure</structure:Name>
    </structure:Concept>
  </structure:ConceptScheme>
  <structure:ConceptScheme id="CROSS_DOMAIN" version="1.0" agencyID="IT1" isFinal="false">
    <structure:Name xml:lang="it">Concetti generali</structure:Name>
    <structure:Name xml:lang="en">Cross domain concepts</structure:Name>
    <structure:Concept id="ADJUSTMENT">
      <structure:Name xml:lang="it">Correzione</structure:Name>
      <structure:Name xml:lang="en">Adjustment</structure:Name>
      <structure:Description xml:lang="it">The set of procedures employed to modify statistical data to enable it to con
text: Adjustments may be associated with changes in definitions, exchange rates, prices, seasons and other factors. Adjustments a
tment can be distinguished from editing and imputation, in that before adjustment, the data are already of sufficient quality t
l.</structure:Description>
    </structure:Concept>
  </structure:ConceptScheme>

```

Fig.5: dettaglio della stringa della DSD contenente un esempio di ConceptScheme con i relativi concetti

3. Query sui dati

Sottoponendo al metodo "GetCompactData" del web-service la query "SDMX_Compact_Data_Query_Example.xml" in cui, nel tag <query:Dataflow> è stato inserito il codice identificativo (id) del dataflow di interesse (ottenuto tramite la query illustrata nel punto 2) è possibile estrarre i dati relativi ad un determinato dataflow in formato SDMX Compact.

```

<Query>
  <query:DataWhere>
    <query:And>
      <query:Dataflow>114_196</query:Dataflow>
    </query:And>
  </query:DataWhere>
</Query>

```

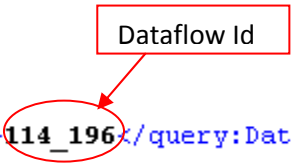


Fig.6: Dettaglio relativo alla query "SDMX_Compact_Data_Query_Example.xml" con l'indicazione del tag in cui inserire l'id del dataflow di interesse

```

<nsi:DataSet>
  <nsi:Series FREQ="M" REF_AREA="IT" ADJUSTMENT="N" INDICATOR="IND_PROD" ACTIVITY="0020" BASE_PER="2005" UN_MEAS="4">
    <nsi:Obs TIME_PERIOD="1990-01" OBS_VALUE="91.5" />
    <nsi:Obs TIME_PERIOD="1990-02" OBS_VALUE="92" />
    <nsi:Obs TIME_PERIOD="1990-03" OBS_VALUE="101.8" />
    <nsi:Obs TIME_PERIOD="1990-04" OBS_VALUE="88.7" />
    <nsi:Obs TIME_PERIOD="1990-05" OBS_VALUE="99.7" />
    <nsi:Obs TIME_PERIOD="1990-06" OBS_VALUE="95.6" />
    <nsi:Obs TIME_PERIOD="1990-07" OBS_VALUE="98.5" />
    <nsi:Obs TIME_PERIOD="1990-08" OBS_VALUE="46.2" />
    <nsi:Obs TIME_PERIOD="1990-09" OBS_VALUE="93.6" />
    <nsi:Obs TIME_PERIOD="1990-10" OBS_VALUE="102.9" />
    <nsi:Obs TIME_PERIOD="1990-11" OBS_VALUE="94.5" />
    <nsi:Obs TIME_PERIOD="1990-12" OBS_VALUE="79.7" />
  </nsi:Series>
</nsi:DataSet>

```

Fig.7: Dettaglio relativo alla stringa di dati in formato compact relativa alla query "SDMX_Compact_Data_Query_Example.xml"

Nella stessa maniera si può sottoporre al metodo "GetGenericData" del web-service la query "SDMX_Generic_Data_Query_Example.xml" in cui, nel tag <query:Dataflow> è stato inserito il codice identificativo (id) del dataflow di interesse (ottenuto tramite la query illustrata nel punto 2) è possibile estrarre i dati relativi ad un determinato dataflow in formato SDMX Generic.

E' anche possibile impostare filtri (SDMX constraints) nell query di dati, basati sui codici di deminsioni ed attributi posti in 'AND' O IN 'OR', come riportato nell'esempio seguente:

```

<Query>
  <query:DataWhere>
    <query:And>
      <query:Time>
        <query:StartTime>2011-01</query:StartTime>
        <query:EndTime>2012-12</query:EndTime>
      </query:Time>
      <query>Dataflow>114_196</query>Dataflow>
      <query:And>
        <query:Or>
          <query:Dimension id="ADJUSTMENT">N</query:Dimension>
        </query:Or>
        <query:Or>
          <query:Dimension id="ACTIVITY2007">256</query:Dimension>
          <query:Dimension id="ACTIVITY2007">259</query:Dimension>
        </query:Or>
      </query:And>
    </query:And>
  </query:DataWhere>

```

Fig.8: Esempio di filetro o constraint posto sulle query di dati

Esempi di query di dati filtrate sono riportati nei seguenti file:" *SDMX_Compact_Filtered_Data_Query_Example.xml*" ed "*SDMX_Generic_Filtered_Data_Query_Example.xml*"

Come si può osservare, è possibile, attraverso gli step descritti e la corretta identificazione dei parametri, impostare una query che restituisca i dati desiderati. Nella figura 7 in cui è rappresentato un esempio di messaggio di dati in formato "compact", è possibile vedere che gli attributi XML del tag "Series" contengono i valori delle dimensioni le cui combinazioni identificano in modo univoco la serie storica, mentre i tag "Obs", figli di "Series", contengono ciascuno l'indicazione dei periodi temporali relativi alle varie istanze della serie (attributo xml "TIME_PERIOD") ed i corrispondenti valori osservati (attributo xml "OBS_VALUE ")

Appendice 1 - SDMX (Statistical Data and Metadata eXchange)

SDMX (Statistical Data and Metadata eXchange) nasce dalla collaborazione di alcune tra le più importanti organizzazioni internazionali con lo scopo di favorire e migliorare lo scambio e la condivisione di dati e metadati statistici.

Lo standard SDMX è composto essenzialmente da:

- un Information Model (IM), dal quale tutti gli oggetti informativi e i messaggi di scambio vengono derivati;
- le Content Oriented Guidelines (COG) che rappresentano un insieme di raccomandazioni per descrivere i dati statistici in maniera armonizzata, “cross-domain” e “cross-organization”;
- una Architettura per permettere un efficiente scambio e condivisione di dati e metadati.

In particolare l’SDMX Information Model (SDMX- IM) consiste di un insieme di oggetti funzionali, organizzati, nella versione 2.0 dello standard, su tre livelli concettuali:

I - SDMX Base

II - Structural Definitions

III - Reporting and Dissemination.

Per descrivere e documentare i dati, lo standard nella sua versione 2.0 fa riferimento a entità (in SDMX “artefatti”) organizzate in modo tale da rappresentare i dati e referenziarli adeguatamente:

Id, agency e version sono i tre elementi identificativi di un artefatto SDMX.

Id è il codice identificativo dell’ artefatto.

Agency è dato dal nome dell’organizzazione che è creatrice e/o proprietaria dell’ artefatto.

Version fornisce la versione dell’ artefatto. In particolare, se l’ artefatto è finalizzato, non è possibile modificarlo a meno di non creare una versione nuova dello stesso.

I principali artefatti atti a descrivere i dati e i metadati statistici sono:

Data Flow: struttura che descrive il contenuto di un set di dati che l’organizzazione produttrice fornisce per differenti periodi temporali di riferimento. L’elemento caratterizzante dei Data Flow, è la Key Family in quanto ne definisce la struttura costitutiva in termini di componenti(dimensioni, attributi, misure).

Key Family o Structure Definition (DSD) è la definizione di una struttura dati Data in termini dei suoi componenti, ovvero di dimensioni, attributi e misure.

Le dimensioni : ovvero componenti che identificano e descrivono il fenomeno osservato.

Gli attributi : ovvero componenti che descrivono solamente, il fenomeno osservato.

Le misure : componenti che rappresentano il fenomeno o i fenomeni osservati.

Concept Scheme: è un raggruppamento di concetti che referenziano le componenti(dimensioni,attributi,misure) della Key Family.

Code List: è una lista di codici associati a componenti (dimensioni e attributi codificati) delle DSD.

Category Scheme: è un insieme di categorie, organizzate gerarchicamente, che classificano i Data Flow.

La versione 2.0 dello standard SDMX ammette tre formati di messaggi per lo scambio dei dati, di seguito dettagliati.

Generic data Message: è l'unico messaggio di dati indipendente da una Data Structure Definition (DSD) perché la struttura dei dati è insita nel messaggio. E' usato quando le applicazioni che ricevono i dati non hanno conoscenza della struttura dati prima di ricevere i dati stessi.

Compact data Message: permette di scambiare e condividere grandi moli di dati in quanto risulta molto sintetico rispetto al formato precedente, viene usato per condividere e scambiare dati in serie storiche.

Cross Sectional Message: permette di trattare più fenomeni osservati nello stesso dataset .