SDDS Import

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of Document** | Project design | | | | |
| **Reference:** |  | | | | |
| **Issue:** |  | **Revision:** |  | **Status:** | Draft |
| **Created by:** | Massimiliano Gianni e Proietti Valerio | | | **Date:** | 19/09/2014 |
| **Updated by:** | Proietti Valerio | | |  | 26/09/2014 |
| **Approved by:** |  | | | | |

Sommario

[Overview 3](#_Toc399411923)

[Main functional requirements 4](#_Toc399411924)

[Main not-functional requirement 5](#_Toc399411925)

[Possibili implementazioni 6](#_Toc399411926)

[Amministrazione 7](#_Toc399411927)

[Gestione connessione DB 7](#_Toc399411928)

[Gestione Utenti 7](#_Toc399411929)

[Upload artefatti 8](#_Toc399411930)

[Upload del file 8](#_Toc399411931)

[Selezione artefatti da importare 8](#_Toc399411932)

[Importazione 8](#_Toc399411933)

[Gestione dimensioni 9](#_Toc399411934)

[Selezione attributi conditional 9](#_Toc399411935)

[Creazione dimensioni 9](#_Toc399411936)

[Caricamento dati con file CSV 10](#_Toc399411937)

[Selezione DataFlow 10](#_Toc399411938)

[Upload CSV 11](#_Toc399411939)

[Selezione Mapping 12](#_Toc399411940)

[Immissione dati 13](#_Toc399411941)

[Log 14](#_Toc399411942)

[Configurazione Log 14](#_Toc399411943)

[Database 15](#_Toc399411944)

[DB artefatti SDMX 15](#_Toc399411945)

[DB di Datawarehouse 15](#_Toc399411946)

[DB di Amministrazione 15](#_Toc399411947)

[DB di Mapping 16](#_Toc399411948)

# Overview

Il progetto presentato richiede la costruzione di un’applicazione Web che permetta tramite Web GUI ad un utente l’inserimento di dati in database relazionali per mezzo di file CSV e di strutture dati a partire da file di struttura contenenti metadati strutturali (arterfatti) SDMX.

Il comma-separated values (abbreviato in CSV) è un formato di file basato su file di testo utilizzato per l'importazione ed esportazione (ad esempio da fogli elettronici o database) di una tabella di dati. In questo formato, ogni riga della tabella (o record della base dati) è normalmente rappresentata da una linea di testo, che a sua volta è divisa in campi (le singole colonne) separati da un apposito carattere separatore, ciascuno dei quali rappresenta un valore.

L’utente nel momento del caricamento dei dati dovrà avere a disposizione un dataflow (SDMX dataset) precedentemente creato, sarà quindi possibile all’utente oltre l’inserimento dei dati, anche la creazione di dataflow (con la definizione della relativa struttura) con apposita Web GUI.

# Main functional requirements

|  |  |
| --- | --- |
| FR1-A | Amministrazione:   1. Gestione connessione DB    1. Creazione schema DB 2. Gestione utenti (gestione autenticazione e autorizzazioni). |
| FR2-A | Upload artefatti:   1. Upload del file SDMX 2. Selezione artefatti da importare 3. Importazione |
| FR3-A | Creazione dataset ed inserimento dimensioni:   1. Definizione struttura tabelle dei fatti (DF) a partire dalle DSD SDMX (selezione attributi conditional, dimensioni e attributi mandatory obligatori). 2. Creazione tabella dei fatti e popolamento a partire dalle codelist SDMX. |
| FR4-A | Creazione Mapping:   1. Creare relazioni tra colonne del CSV e tabella dei fatti (DF). 2. Conservare Mappature per poterle sfruttare successivamente. |
| FR5-A | Caricamento dati:   1. Selezione e creazione tabelle dei fatti, se utente abilitato (vedi punto FR3-A2) 2. Gestire upload di file CSV (anche molto grandi > 1gb).    1. Selezione file.    2. Preview    3. Caricamento    4. Notifica. |

# 

# Main not-functional requirement

|  |  |
| --- | --- |
| NFR1-A | L’applicazione dovrà connettersi a database SQL Server v 2012 (Oracle, mySQL prossime implementazioni). |
| NFR2-A | Possibilità di salvare il file XML mapping in Directory associata all’utente o su db. |
| NFR3-A | Ogni utente è abilitato all’inserimento solo in domini statistici assegnategli. |

# Possibili implementazioni

L’ambiente di sviluppo scelto è Visual Studio 2010, il linguaggio ASPX con C# .NET, il database Microsoft SQL Server con linguaggio T-SQL.

Sarà previsto l’utilizzo di tecniche Ajax per l’upload dei file di grandi dimensioni.

Le WebGUI saranno implementate in javascript con l’utilizzo del framework jQuery ed eventuali Plugin di esso.

# Amministrazione

## Gestione connessione DB

Le utenze con profilo di Amministratore potranno configurare la connessione verso il DB utilizzata dall'applicazione . Sarà possibile impostare una connessione verso un DB esistente vuoto, o verso un DB popolato. Nel primo caso verranno creati gli oggetti del DB(tabelle, viste, stored procedure ecc...), mentre nel secondo caso verrà solamente impostata la connessione verso il DB, senza applicare nessun tipo di trasformazione degli oggetti.

## Gestione Utenti

Gli utenti sono divisi in **amministratore** e **operatore.**

Le informazioni che verranno associate ad ogni utente saranno:

Username, Password, Ruolo, Nome, Cognome, Email (per eventuali notifiche),Agenzia, flag abilitato a creare definire DF.

Un amministratore può:

1. Creare, eliminare e aggiornare utenti.
2. Assegnare DF ad utenti.
3. Definire connessioni.
4. Caricare metadati.
5. Creare DF.
6. Creare Mapping.
7. Caricare dati.

L’operatore può:

1. Creare DF, se abilitato.
2. Creare Mapping dati.
3. Caricare Dati.

# 

# Upload artefatti

Questa funzionalità permette di importare file di artefatti(DSD,Codelist, Concept scheme) nel sistema.

## Upload del file (Strutturali SDMX)

Le utenze con profilo di Amministratore potranno importare file di artefatti nel sistema. Tramite un controllo di tipo Uploadfile, l'utente potrà selezionare il file da importare. Premendo il tasto "Importa" il file verrà copiato in una directory di appoggio dell'applicazione(server side).

## Selezione artefatti da importare

Terminato l'upload del file, verranno mostrati a video gli artefatti importabili(DSD,Codelist, Concept scheme). Se nel file saranno presenti artefatti non importabili non saranno visualizzati.

Tramite una lista a selezione sarà possibile selezionare e scegliere gli artefatti da importare sul DB.

## Importazione

Selezionati gli artefatti da importare, tramite la pressione del tasto "Importa", sarà possibile procedere all'importazione su DB. Tramite il richiamo di apposite procedure C# e SQL, verranno create le tabelle delle dimensioni....

# Creazione dataflow (dataset) e caricamento dati con file CSV

Nella creazione di un dataflow la procedura sarà:

1. Lettura DSD + CL + CS memorizzate nel database.
2. Selezione di un DSD.
3. Selezione campi chiave e attributi condizionali (ove le dimensioni risulteranno minori di 16 saranno selezionate tutte senza possibilità di deselezionarle altrimenti se maggiori di 16 dovranno essere scelte).
4. Creazione tabelle dei fatti (FILTS\_, FACTS\_).
5. Creazione tabelle Dimensioni (DIMM\_).
6. Popolamento tabelle Dimensioni con tutti gli item disponibili.

Nell’attività di caricamenti dati un OP procederà in sequenza:

## Selezione DataFlow

Si deve selezionare un dataflow per ottene una lista di dimensioni. Quindi verrà visualizzata una lista dei Dataflow accessibili i quali potranno essere stati utilizzati oppure no.

## Upload CSV

Procederà alla selezione del file CSV tramite controllo UploadFile:

Il file csv sarà composto da una matrice di valori, separati tra loro con un carattere di separazione (colonne), e per riga con il carattere di fine riga righe.

Quindi un esempio di file CSV potrebbe essere:

0040|N|IT|IND\_PROD|M|1994M4|95,8

0040|N|IT|IND\_PROD|M|1995M3|115,5

0040|N|IT|IND\_PROD|M|1996M2|103,8

0040|N|IT|IND\_PROD|M|1996M3|107,4

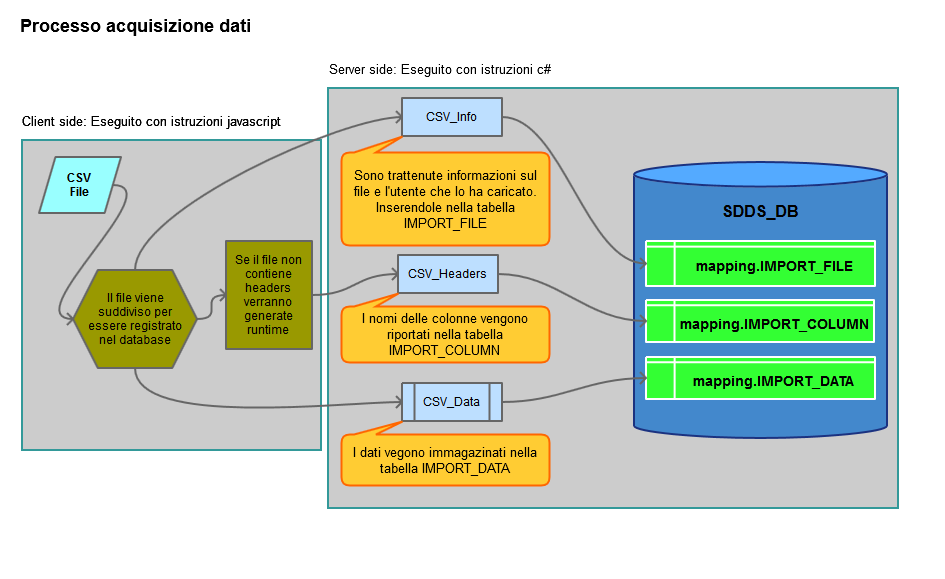
0040|N|IT|IND\_PROD|M|1996M12|78,7

0040|N|IT|IND\_PROD|M|1997M11|107,1

Ove nella prima riga non siano presenti nomi di colonna, saranno generati automaticamente come : CLN\_1,CLN\_2,CLN\_3,CLN\_4,CLN\_5,CLN\_6,CLN\_7 per gestire tutte le possibili situazioni senza allontanarsi del processo iniziale di mapping.

Dal contenuto del file saranno estrapolate informazioni relative ad: FileInfo(filename, size), CSVHeaders (lista dei nomi delle colonne), CSVData (matrice di valori).

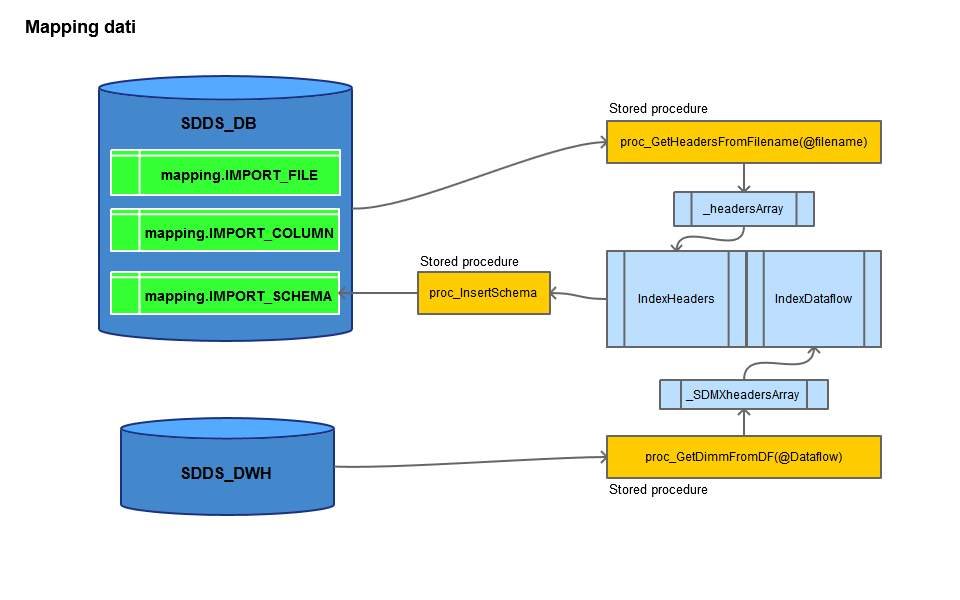
I dati verranno inseriti nel DB, opportunamente distribuiti in tabelle.



## Selezione Mapping

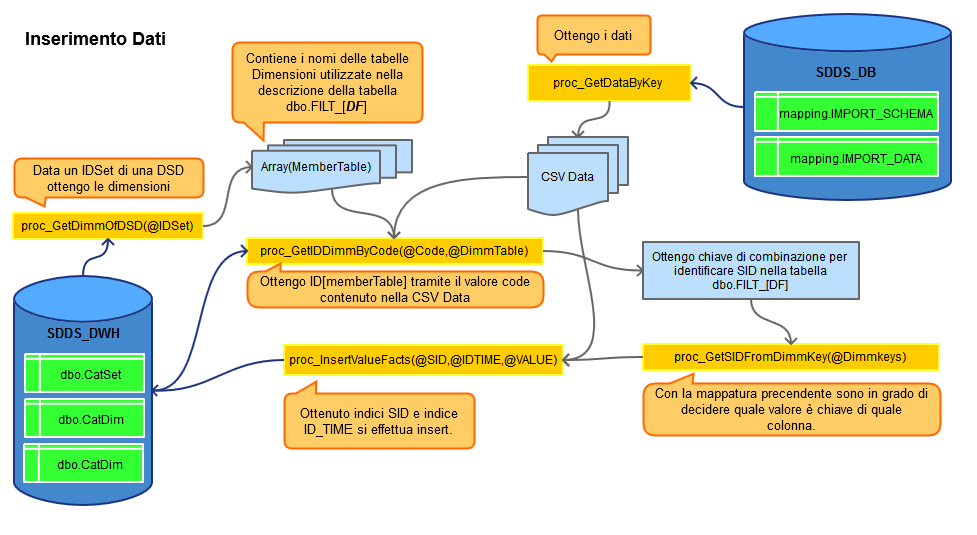
Visualizzati i mapping dati utilizzati precedentemente l’operatore potrà procedere con:

* 1. Selezione di un Mapping dati esistente (se idoneo).
  2. Creazione nuovo mapping dati:
     1. Il sistema già avrà ottenuto i CSVHeaders e li presenterà in un controllo di tipo ListView selezionabile.
     2. Sarà presentata affianco alla ListView degli headers CSV una lista sempre con controllo ListView contenente le dimensioni nel DF scelto.
     3. Tramite controlli di tipo Button l’operatore procederà a creare relazioni di tipo 1 -1 tra gli HeadersDSV e DimensionDF.
     4. Tramite controllo Button potrà salvare il mapping dati nel database, potendolo riutilizzare in futuro su file CSV con medesimi HeadersCSV.
     5. Il file verrà nominato con pattern: nome utente\_nomedataflow\_datetimenow



## Immissione dati

1. Avvierà l’inserimento dei dati .
   * 1. L’inserimento dei dati avverrà server side con esecuzione di query SQL.
     2. A termine procedura il sistema presenterà opportuna notifica.



Una possibile costruzione della Web Gui presenterà i seguenti controlli:

# D:\SDMX_RI\OpenProject\SDDS_Import\doc\webgui.png

# Log

E’ previsto un sistema di log che permetta il monitoraggio delle attività svolte dell’applicazione.

Verrà creato un file di tipo testuale con estensione “.log” dove in modo sequenziale saranno inserite le varie attività svolte.

## Configurazione Log

Il sistema di Log è configurabile tramite file XML che ne descriverà le modalità.

Nel file potranno essere specificate informazioni per:

1. Il percorso fisico su disco della directory contente i file .log .
2. Il livello di dettaglio del log (ALL, DEBUG, INFO, WARNING, ERROR, NONE).

La libreria di che implementa le funzionalità di Log scelta sarà log4net, ampiamente testata ed utilizzata all’interno delle applicazioni ISTAT.

# Database

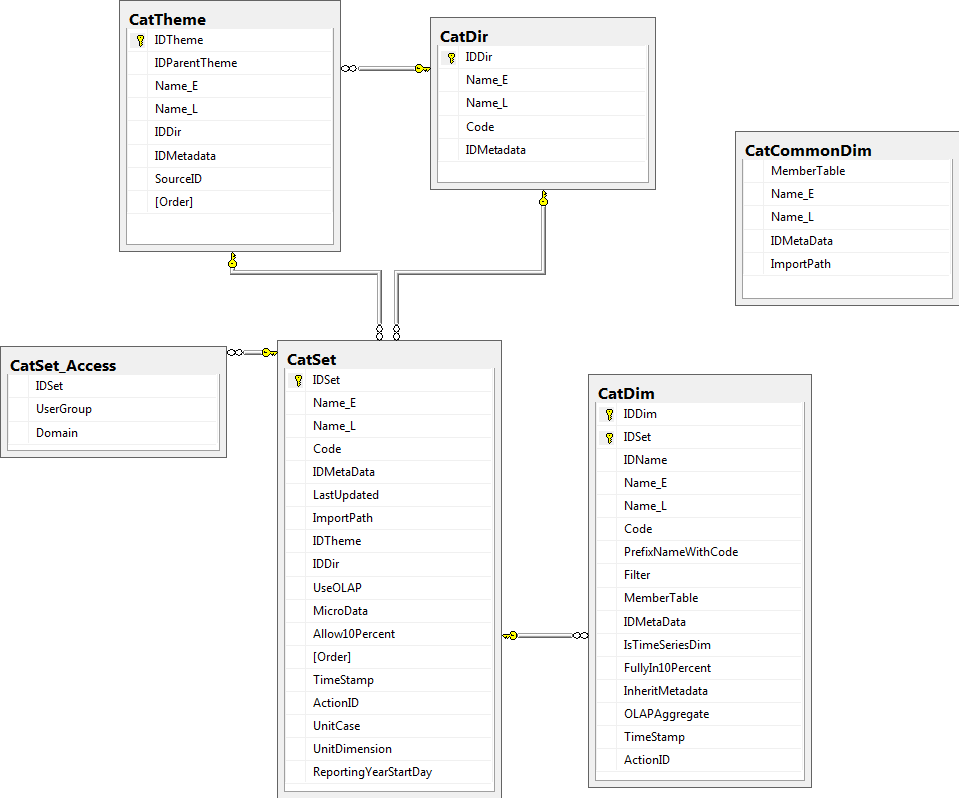
Vengono di seguito descritte le varie sezioni del DB utilizzate dall'applicazione.

## DB schema artefatti SDMX

In questa porzione del DB sono presenti le tabelle relative all'importazione delle strutture SDMX....

## DB schema di Datawarehouse

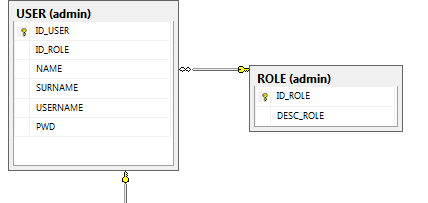
In questa porzione del DB sono presenti le tabelle relative al Datawarehouse.



## 

## DB schema di Amministrazione

In questa porzione del DB sono presenti le tabelle di amministrazione.



## DB schema di Mapping

In questa porzione del DB sono presenti le tabelle relative al mapping dei dati tra la parte SDMX e di Datawarehouse.

La Web Application permette la memorizzazione dei dati e degli schemi di mapping attraverso tabelle su db.

Per ogni file vengono memorizzate le colonne e per ogni colonna sono conservati i valori.

In aggiunta per ogni colonna di un determinato file viene memorizzata la corrispettiva tabella di dimensione contenuta nel db Datawarehouse.

