



*Agenzia Nazionale per i Servizi Sanitari Regionali*

# Manuale theMatrix/Neo



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## Sommario

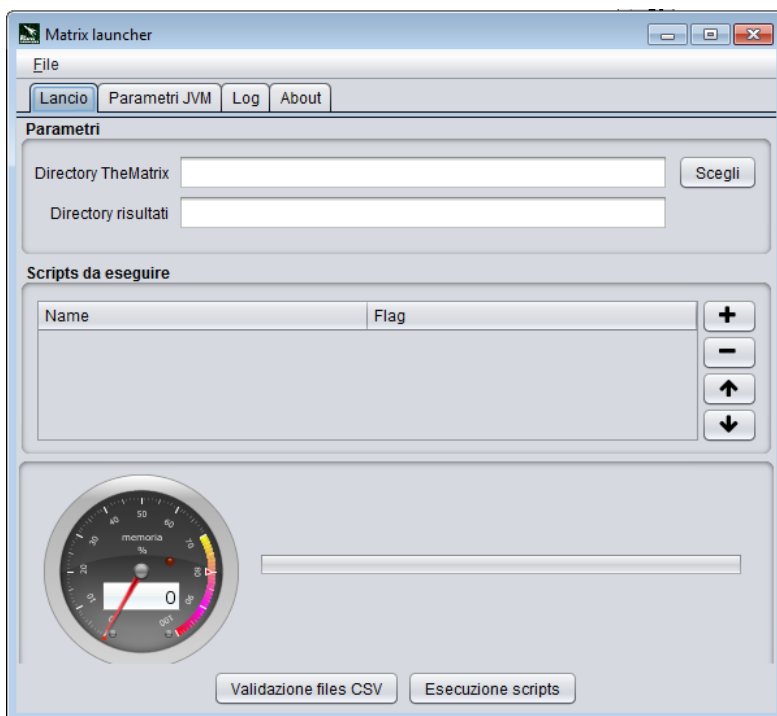
1	Installazione.....	3
2	Caricamento dei dati .....	5
2.1	Formato dati .....	5
2.2	Dati di test .....	5
2.3	Dati reali .....	5
2.4	Generazione xml.....	6
2.5	Pulizia dei dati.....	6
3	Lancio di script di qualità.....	8
4	Operazioni ordinarie e di mantenimento.....	8
4.1	Eseguire script .....	8
4.2	Aggiornare i dati .....	10
4.3	Sviluppare script .....	10
5	Documentazione.....	11
6	Allegato 1.....	12



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## 1 Installazione

TheMatrix funziona con sistema operativo Windows (dalla versione 7 in poi), Unix o Mac OS, e richiede che sia installato Java 7 o superiori.



Per installare TheMatrix agisci come segue

1. scarica il file dal link di Download di [questa pagina](#)
2. esegilo. Durante l'installazione ti verrà chiesto
  - se vuoi le interfacce in inglese o italiano
  - se accetti la licenza (GPL v3)
  - in quale directory vuoi che TheMatrix sia installato (si raccomanda di scegliere una directory vuota)

TheMatrix è ora installato. Sul desktop trovi due link



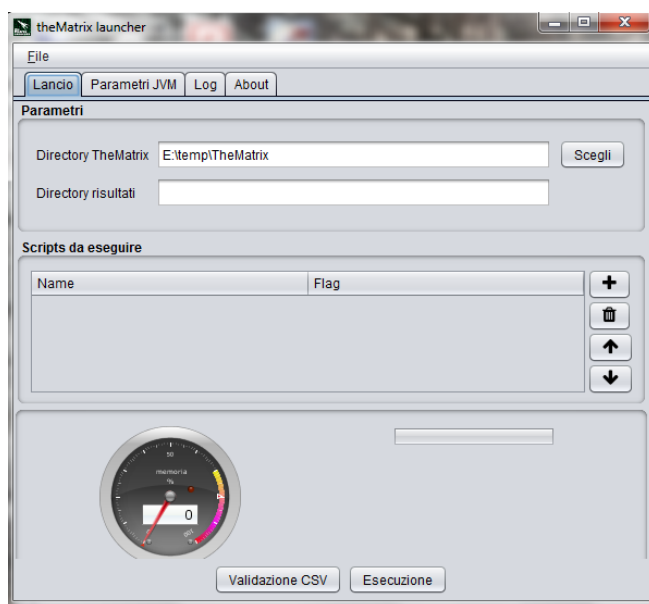
questo è il link a TheMatrixScriptWriter, l'interfaccia per *sviluppare* gli script; in questo documento *non* ci occupiamo di questo tool.



questo è il link a TheMatrixLauncher, l'interfaccia per *eseguire* gli script. Lanciando questo link si apre questa interfaccia



# *Agenzia Nazionale per i Servizi Sanitari Regionali*



Il folder dell'installazione contiene

- il programma TheMatrixLauncher
- la directory TheMatrixScriptWriter
- la directory TheMatrix, che contiene il motore del programma, e in particolare
  - la directory TheMatrix/iad, che contiene i file di input in formato csv (ci sono due piccoli file di test, puoi aprirli con un editor di testo o con un foglio di calcolo)
  - la directory TheMatrix/scripts, che conterrà gli script (ora è vuota)
  - la directory TheMatrix/results, che conterrà i risultati degli script (ora è vuota)

Per far funzionare TheMatrix è necessario caricare nella directory TheMatrix\iad i dati, in particolare quelli dei flussi amministrativi, e caricare nella directory TheMatrix\scripts gli script che contengono le istruzioni.



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## 2 Caricamento dei dati

### 2.1 Formato dati

I dati amministrativi devono essere esportati nel formato IAD descritto nell'appendice di questo documento, e salvati nella cartella TheMatrix\iad. I dati sono estratti dalle 6 fonti principali

PERSON - l'anagrafe

EXE - le esenzioni

HOSP - le schede di dimissione ospedaliera

DRUGS - la farmaceutica territoriale

DDRUG - i farmaci a erogazione diretta

OUTPAT - le prestazioni ambulatoriali

### 2.2 Dati di test

Se si desidera eseguire un test prima di scaricare i propri dati reali, utilizzare i dati di esempio contenuti a [questo link](#): si tratta di dati falsi di una ipotetica ASL di 5000 assistiti. Per proseguire nel test, seguire le istruzioni contenute nelle sezioni 2.4, 2.5 e 3 di questo documento. Si consiglia di farlo prima di caricare i propri dati reali.

**Una volta completato il test, riprendere dal punto 2.3  
(qui di seguito)**

### 2.3 Dati reali

Nei dati di esempio contenuti a [questo link](#) ogni flusso è contenuto in un unico file csv. Se si ritiene preferibile, però, quando si formattano i propri dati reali, i dataset EXE, HOSP, DRUGS, DDRUG e OUTPAT possono essere creati a pezzi, purché i file csv corrispondenti abbiano nomi che cominciano con il nome standard del flusso. Per esempio, se si vogliono suddividere i dati di DRUGS (la farmaceutica territoriale) in diversi file, bisogna salvare i dati nella directory TheMatrix\iad in file csv i cui nomi che cominciano con "DRUGS", per esempio con nomi di questo tipo

DRUGS2006\_7.csv

DRUGS2008\_12.csv

DRUGS2013.csv

Anche PERSON può essere creata in più file, ma la logica in questo caso è diversa: si possono utilizzare [questi suggerimenti](#).

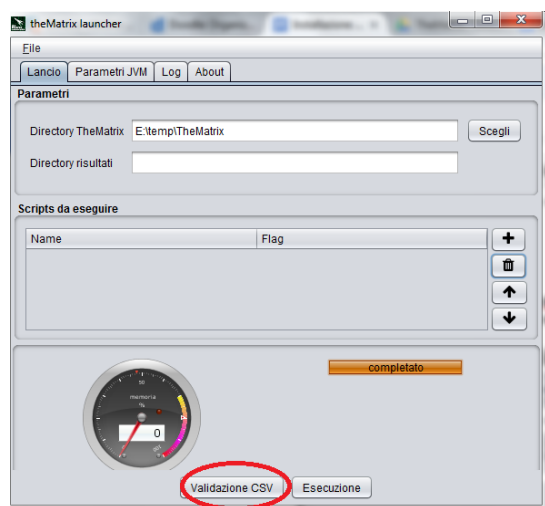
**ATTENZIONE:** non usare trattini "-" nei nomi di questi file.



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## 2.4 Generazione xml

Per rendere visibili i file csv a TheMatrix è necessario validarli generando un piccolo file xml per ciascuno. Per fare questo, lanciare TheMatrixLauncher e premere il tasto 'Validazione CSV'. Per file molto grossi l'operazione può richiedere qualche minuto.



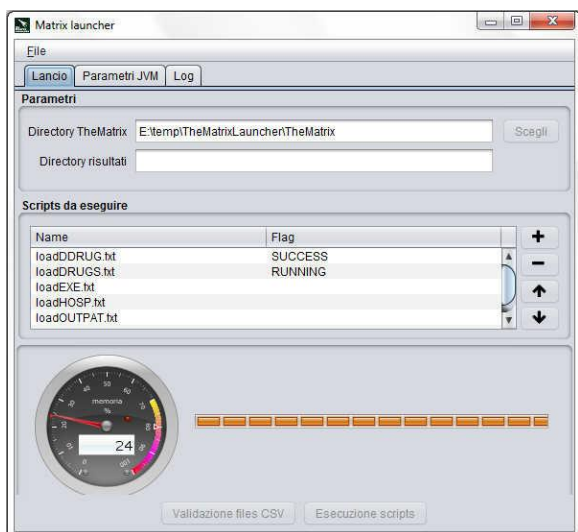
## 2.5 Pulizia dei dati

Prima di avviare gli script ordinari di TheMatrix è necessario ripulire i file di IAD. Per eseguire questa operazione vanno lanciati alcuni script speciali chiamati *script di load*. Agire come segue:

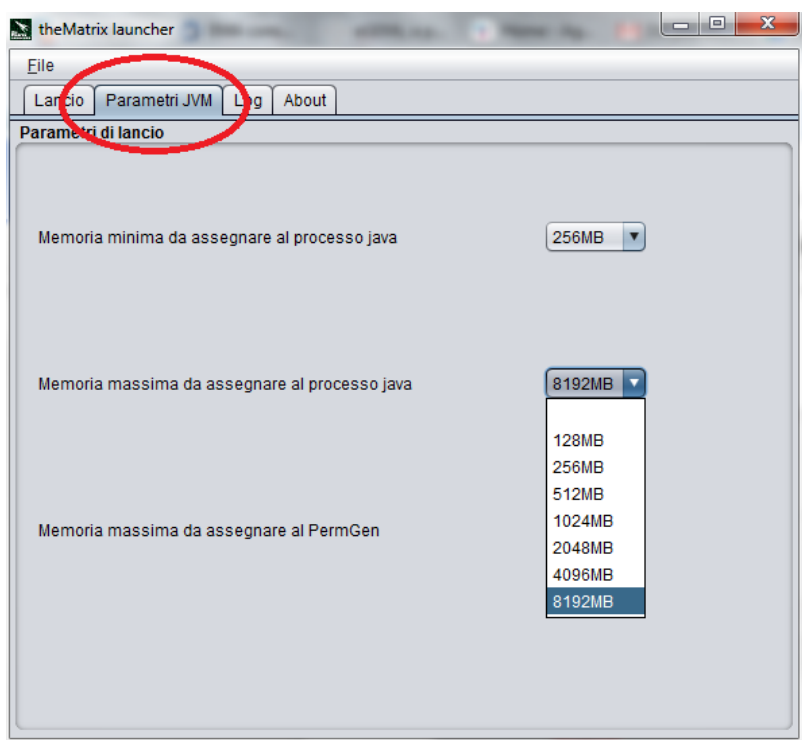
1. scaricare nella directory TheMatrix\lookups [questo file](#) e decomprimerlo
2. scaricare nella directory TheMatrix\scripts i file a [questo link](#)
3. dall'interfaccia premere il tasto +: compare la lista degli script. Selezionare tutti e 6 gli script appena scaricati: cominciano tutti con la stringa "load"
4. Premere il tasto 'Esecuzione scripti'. L'esecuzione si avvia. Accanto allo script in corso di esecuzione compare la scritta 'RUNNING'. accanto agli script completati con successo compare la scritta 'SUCCESS'. Se si desidera, i messaggi generati durante l'esecuzione possono essere seguiti nel tab 'log'. La sequenza può impiegare qualche minuto, fino a qualche ora prima di essere completata (per esempio sui dati dell'intera regione Toscana, che sono contenuti in file della dimensione complessiva di circa 100G, impiega circa 14 ore)



# *Agenzia Nazionale per i Servizi Sanitari Regionali*



**Nota.** Si consiglia, in caso si gestiscano dataset molto grandi e si abbia a disposizione un calcolatore sufficientemente potente, di accedere al tab 'Parametri JVM' di TheMatrixLauncher e aumentare le risorse messe a disposizione del programma, utilizzando i tre menù a tendina (si veda la figura qui sotto)





# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## 3 Lancio di script di qualità

Per verificare che i dati siano interpretati correttamente da TheMatrix procedere come segue


1. scaricare nella directory TheMatrix\scripts i file di [questo link](#)
2. dall'interfaccia premere il tasto +: compare la lista degli script. Selezionare tutti e 6 gli script appena scaricati: cominciano tutti con la stringa "qualita"
3. Premere il tasto 'Esecuzione scripts'. L'esecuzione si avvia, e può essere seguita dal tab 'log'. Altrimenti il tab principale elenca quali script sono stati eseguiti e quale script è in via di esecuzione. La sequenza può impiegare qualche minuto, fino a qualche ora prima di essere completata (per esempio sui dati dell'intera regione Toscana, che sono contenuti in file della dimensione complessiva di circa 100G, impiega circa 3 ore)
4. Alla fine di ciascuno script nella directory TheMatrix\results compare un piccolo file di nome QualitaYYYYYXXXX.csv, in cui XXXX è un anno (ad esempio 2014) e YYYYY è una stringa tra PERSON, EXE, HOSP, DDRUG, DRUGS e OUTPAT. Aprirlo e
  - a. nel caso di PERSON: il file riporta il numero di righe del file PERSON e il numero di soggetti che risultano assistiti nel 2005, 2006,... fino all'anno XXXX: verificare che i dati siano corretti
  - b. negli altri casi: il file riporta il numero di assistiti al 1 gennaio XXXX, il numero di righe del file YYYYY riferite all'anno XXXX e a un soggetto assistito al 1 gennaio XXXX, e il numero medio di righe per soggetto. Confrontare questi dati con quelli prodotti dalla ASL/regione con le sue procedure consuete

Se i dati risultano coerenti è possibile a questo punto eseguire qualsiasi script.

## 4 Operazioni ordinarie e di mantenimento

### 4.1 Eseguire script

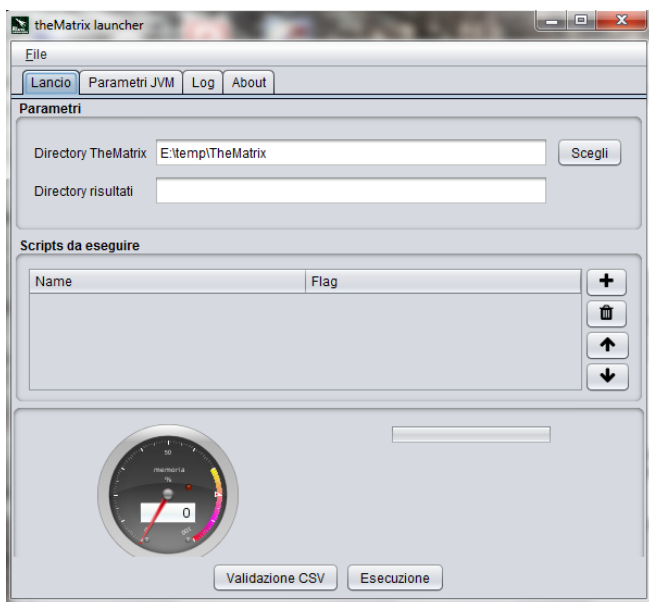


1. Lanciare TheMatrixLauncher usando il link  del desktop oppure l'eseguibile contenuto nella directory di installazione: questo causerà l'apparizione della seguente interfaccia

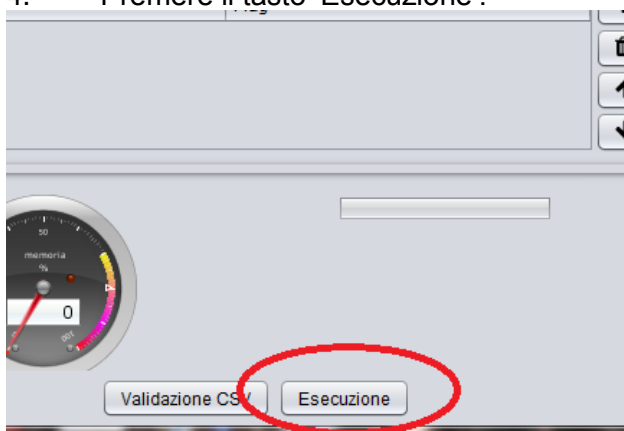




# *Agenzia Nazionale per i Servizi Sanitari Regionali*



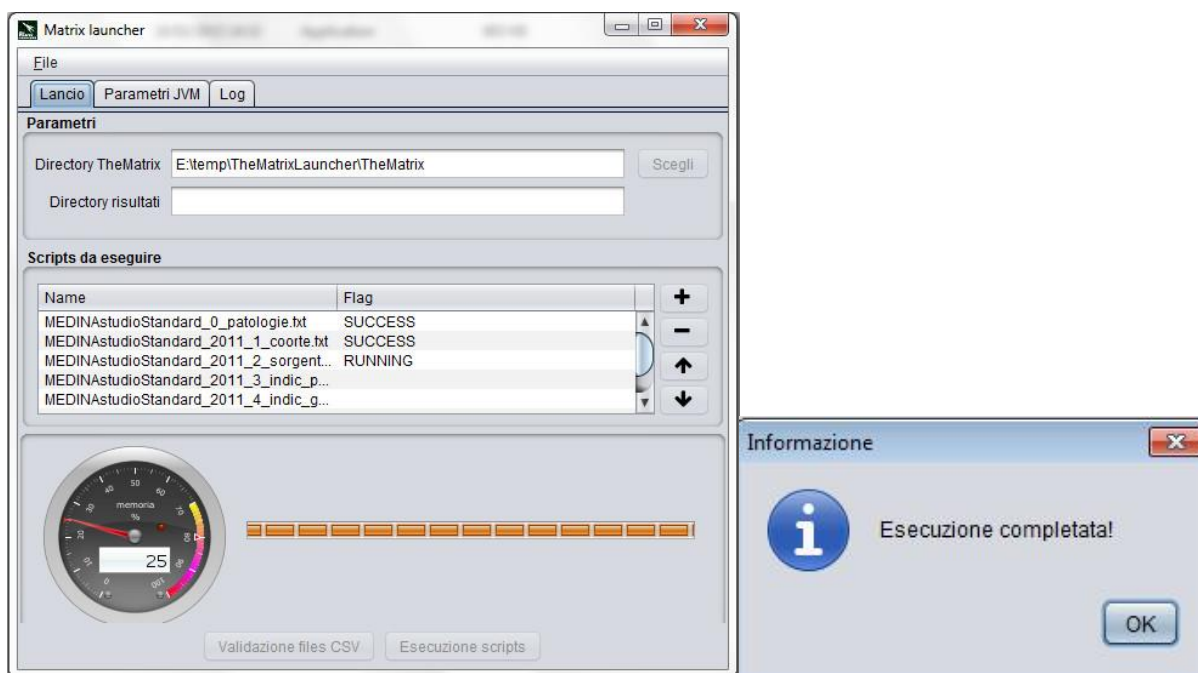
2. Salvare nella directory TheMatrix\scripts gli script da eseguire
3. Dall'interfaccia premere il tasto +: compare la lista degli script. Selezionare gli script da eseguire
4. Premere il tasto 'Esecuzione'.



5. L'esecuzione si avvia. Accanto allo script in corso di esecuzione compare la scritta 'RUNNING'. Accanto agli script completati con successo compare la scritta 'SUCCESS' (si veda la figura sottostante).



# *Agenzia Nazionale per i Servizi Sanitari Regionali*



6. L'esecuzione può impiegare da qualche minuto fino a qualche ora prima di essere completata: dipende dalla complessità della procedura, dalla dimensione dei file trattati e dalle caratteristiche del computer su cui è installato TheMatrix. Al termine dell'esecuzione compare questa finestra. I file prodotti dalla procedura si trovano nella directory TheMatrix\results

## 4.2 Aggiornare i dati

Se i dati in formato IAD vanno aggiornati, aggiungendo per esempio un nuovo anno, si può scegliere di rifare un caricamento "da zero": cancellare i vecchi csv e xml da TheMatrix\iad, ricreare dei file csv unici contenenti tutti i dati (vecchi e nuovi), caricarli in TheMatrix\iad, e ripetere le procedure dei punti 2.4 e 2.5.

Altrimenti, procedere come segue

1. generare nuovamente PERSON.csv, per tutto il periodo considerato, oppure utilizzare nuovamente [questa strategia](#), e sostituire i file vecchi in TheMatrix\iad con quelli nuovi. Cancellare i file xml corrispondenti
2. generare i file EXE\_nomeanno.csv, DRUGS\_nomeanno.csv, DDRUG\_nomeanno.csv, HOSP\_nomeanno.csv e OUTPAT\_nomeanno.csv e salvarli in TheMatrix\iad, senza toccare i file precedenti
3. cancellare nella directory TheMatrix\results i file PERSONFile.csv, HOSPFile.csv, EXEFile.csv, DRUGSFile.csv, DDRUGFile.csv, OUTPATFile.csv e i corrispondenti file xml
4. premere nell'interfaccia il tasto 'Validazione CSV'
5. scaricare nuovamente nella directory TheMatrix\lookups [questo file](#) (sostituendo il precedente) e lanciare nuovamente i 6 script di load (vedi sezione 2.5 Pulizia dei dati)

## 4.3 Sviluppare script



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

Per sviluppare uno script seguire le indicazioni presentate in [questa pagina](#) del sito di TheMatrix (in lingua inglese).

## 5 Documentazione

TheMatrix è un software disegnato e sviluppato dall'Agenzia Regionale di Sanità della Toscana e dal laboratorio ISTI del Consiglio Nazionale delle Ricerche, nell'ambito del [progetto MATRICE dell'Agenas](#), finanziato dal Ministero della Salute. Il software è rilasciato con licenza GPL v3. Il sito di TheMatrix è [questo](#).

### **Appendice: il formato IAD**

I dati più importanti su cui TheMatrix agisce sono i flussi amministrativi di dati sanitari. Il formato che è stato stabilito nel corso del progetto MATRICE si chiama IAD (Italian Administrative Database) e ricalca i tracciati record contenuti negli allegati tecnici alla normativa nazionale che ha istituito i diversi flussi. Il formato contiene ad oggi otto tabelle

PERSON - l'anagrafe

EXE - le esenzioni

HOSP - le schede di dimissione ospedaliera

DRUGS - la farmaceutica territoriale

DDRUG - i farmaci a erogazione diretta

OUTPAT - le prestazioni ambulatoriali

HOME - l'assistenza domiciliare

RESIDENT - l'assistenza residenziale

Nei file disponibili a [questo link](#) è descritto nel dettaglio il tracciato record di ciascuna tabella, con i riferimenti alla normativa. Da notare che in IAD le tabelle DRUGS e DDRUG hanno lo stesso tracciato record.



# *Agenzia Nazionale per i Servizi Sanitari Regionali*

## 6 Allegato 1

### Approfondimento su come creare PERSON

Tra i file disponibili a questo link

[http://matrice.agenas.it/Content/Document/documentazione\\_IAD\\_per\\_aziende.zip](http://matrice.agenas.it/Content/Document/documentazione_IAD_per_aziende.zip)

si prova in particolare la descrizione del tracciato record di PERSON, ovvero l'anagrafe assistibili.

In questo documento diamo alcuni suggerimenti per creare questo file, che è il più complesso tra quelli che devono essere messi a disposizione di TheMatrix.

Per semplicità consideriamo come variabile dipendente dal tempo solo la variabile GP\_ID, ovvero l'identificativo del medico che assiste il soggetto. Però lo stesso discorso vale per le altre variabili, per esempio la azienda sanitaria di residenza. Inoltre ci concentriamo sulla variabile DATE\_OF\_DEATH.

Pensa per esempio a questi casi

Il **soggetto A**, assistito a partire dal 3 marzo 2001 con il medico M, e tuttora assistito: su PERSON può avere un solo record, come segue

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03		M		

Il **soggetto B**, assistito a partire da 19 ottobre 2012 con il medico M e con assistenza cessata il 10 gennaio 2013: su PERSON può avere un solo record, come segue

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
B	2012-10-19	2013-01-10	M		

Il **soggetto C**, assistito a partire dal 7 giugno 2003 con il medico M e dal 5 maggio 2011 dal medico N, e tuttora assistito: su PERSON deve avere almeno due record, come segue

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
C	2003-06-07	2011-05-04	M		
C	2011-05-05		N		

Il **soggetto D**, assistito a partire dal 22 agosto 2009 con il medico M e assistito fino alla morte, avvenuta il 3 maggio 2010: su PERSON può avere un solo record, in cui ENDDATE e DATE\_OF\_DEATH coincidono, come segue



## *Agenzia Nazionale per i Servizi Sanitari Regionali*

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
D	2009-08-22	2010-05-03	M	2010-05-03	

Riassumendo la tabella PERSON di questi 4 soggetti può essere compilata come segue

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03		M		
B	2012-10-19	2013-01-10	M		
C	2003-06-07	2011-05-04	M		
C	2011-05-05		N		
D	2009-08-22	2010-05-03	M	2010-05-03	

Però questo non è l'unico modo possibile per rappresentare questi 4 soggetti. Se risulta più comodo, tieni presente che per TheMatrix un record solo oppure diversi record consecutivi sono la stessa cosa! Per esempio puoi rappresentare il soggetto A sia con un unico record, come abbiamo detto

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03		M		

sia, per esempio, con due record, uno che ha come ENDDATE il 31 dicembre 2010 e l'altro che ha come STARTDATE il giorno successivo, il 1 gennaio 2011, come segue

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03	2010-12-31	M		
A	2011-01-01		M		

### **Se i dati anagrafici sono raccolti in fotografie**

Se per esempio per costruire PERSON tu devi partire da due fotografie dell'anagrafe, una scattata il 1 gennaio 2011 e l'altra scattata il 1 gennaio 2014, tu ti trovi a dover assemblare queste informazioni

Il 1 gennaio 2011



## *Agenzia Nazionale per i Servizi Sanitari Regionali*

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03		M		
C	2003-06-07		M		
D	2009-08-22	2010-05-03	M	2010-05-03	

Il 1 gennaio 2014

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03		M		
B	2012-10-19	2013-01-10	M		
C	2003-06-07		N		

Nella foto del 2014 il soggetto D, che è morto da tempo, non si trova più, mentre nella foto del 2011 il soggetto B, che non era ancora entrato al 1 gennaio 2011, non c'è. Inoltre nel 2014 soggetto C risulta avere sempre la stessa STARTDATE, ma un medico diverso, senza che sia specificata la data in cui è avvenuto il cambio di medico.

Il modo migliore di assemblare queste informazioni è il seguente

- 1) nella foto del 2011 aggiungi la data ENDDATE 2010-12-31 in tutti i record che hanno ENDDATE missing, così

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03	<b>2010-12-31</b>	M		
C	2003-06-07	<b>2010-12-31</b>	M		
D	2009-08-22	2010-05-03	M	2010-05-03	

- 2) nella foto del 2014 sostituisci la data STARTDATE con 2011-01-01 tutte le volte che STARTDATE è inferiore a 2011-01-01

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	<b>2011-01-01</b>		M		
B	2012-10-19	2013-01-10	M		
C	<b>2011-01-01</b>		N		



## *Agenzia Nazionale per i Servizi Sanitari Regionali*

3) fai l'unione dei due file

PERSON_ID	STARTDATE	ENDDATE	GP_ID	DATE_OF_DEATH	...
A	2001-03-03	<b>2010-31-12</b>	M		
C	2003-06-07	<b>2010-31-12</b>	M		
D	2009-08-22	2010-05-03	M	2010-05-03	
A	<b>2011-01-01</b>		M		
B	2012-10-19	2013-01-10	M		
C	<b>2011-01-01</b>		N		