

# The Matrix – Il software

Rosa Gini

Agenzia regionale di sanità della Toscana  
Dipartimento di Informatica Medica dell'Università Erasmus di Rotterdam



ISS 3-5 Aprile 2013

# Indice

## Introduzione

Cos'è TheMatrix

Il progetto MATRICE

## Caratteristiche e logica di TheMatrix

Obbiettivi e funzionalità

Linguaggio di programmazione e interfaccia grafica

Requisiti e modalità di installazione

## Conclusioni



# Indice

## Introduzione

Cos'è TheMatrix

Il progetto MATRICE

Caratteristiche e logica di TheMatrix

Conclusioni



# Indice

## Introduzione

Cos'è TheMatrix

Il progetto MATRICE

Caratteristiche e logica di TheMatrix

Conclusioni



# TheMatrix è il software open source sviluppato dal progetto MATRICE

Il progetto prevede il disegno, la sperimentazione, la validazione e l'applicazione ad alcuni casi di studio di un software open source di aggregazione e di integrazione dei dati dei flussi sanitari nazionali, per produrre la lettura dei percorsi assistenziali di pazienti con patologie complesse

dal documento di progetto di MATRICE



# Contesto del progetto MATRICE

**Problema** La gestione dei pazienti con patologie croniche richiede un modello di cura centrato sul territorio

**Ostacoli** Non esistono flussi ad hoc, scarsa interoperabilità, limiti di privacy

**Potenzialità** I dati amministrativi sanitari esistenti possono risolvere il problema tramite applicazione sistematica di algoritmi che usano il record linkage

**Idea progettuale** Costruire un'infrastruttura informatica per la produzione di dataset omogenei tra diverse ASL e individuare tramite validazione gli algoritmi migliori da implementare nel software



# Indice

## Introduzione

Cos'è TheMatrix

Il progetto MATRICE

Caratteristiche e logica di TheMatrix

Conclusioni



# Obiettivi generali

... condividere basi informative comuni per arrivare ad un sistema che legga trasversalmente i sistemi sanitari regionali, rispetto a percorsi assistenziali articolati e continuativi. . .

dal documento di progetto di MATRICE



# Obiettivi generali

... utilizzare unicamente flussi informativi già esistenti, per poter osservare la gestione delle patologie complesse nelle diverse Regioni e Province Autonome, sulla base di dati la cui validità è condivisa con esse, attraverso operazioni di trasformazione del dato la cui validità verrà stabilita sulla base di rigorosi criteri scientifici. . .

dal documento di progetto di MATRICE



# Obiettivi generali

... al fine di contribuire ad evidenziare la misura dell'equità nell'erogazione delle cure territoriali ...

dal documento di progetto di MATRICE



# Obiettivi generali

... dotare i vari livelli istituzionali di una base conoscitiva condivisa che supporti la *governance* delle cure territoriali ...

dal documento di progetto di MATRICE



## Partner

- Agenzia nazionale per i servizi sanitari regionali (Agenas)
- Ministero della Salute
- 5 Regioni (6 ASL + 1 Regione)
- ARS Toscana
- Società Italiana di Medicina Generale (SIMG)
- Istituto di Scienze e Tecnologia dell'Informazione del CNR
- Dipartimento di Informatica dell'Università degli Studi di Milano
- Dipartimento di Informatica Medica dell'Università Erasmus di Rotterdam



# Struttura

- P1 - **Patologie-Modelli informativi** Selezione delle patologie, enumerazione di algoritmi esistenti, validazione, generazione algoritmi innovativi, scelta
- P2 - **Sistemi di indicatori** Selezione, calcolo e validazione di indicatori di adesione a percorsi assistenziali
- P3 - **Data Integration** Contesto, sviluppo e sperimentazione del processo di erogazione dati
- P4 - **Sperimentazione** Modelli di analisi dei flussi di dati secondari, sperimentazione della loro utilità pratica nell'organizzazione del servizio sanitario a livello locale e centrale
- P5 - **Management e comunicazione**



# Indice

Introduzione

Caratteristiche e logica di TheMatrix

Obbiettivi e funzionalità

Linguaggio di programmazione e interfaccia grafica

Requisiti e modalità di installazione

Conclusioni



# Indice

Introduzione

Caratteristiche e logica di TheMatrix

Obbiettivi e funzionalità

Linguaggio di programmazione e interfaccia grafica

Requisiti e modalità di installazione

Conclusioni



# Una risorsa per tutto il sistema

**Database virtuale** TheMatrix realizza un database virtuale con uno specifico linguaggio di interrogazione

**Open source** Il software verrà rilasciato da Agenas con una licenza open source

**Utilizzabile** Sarà quindi utilizzabile da qualsiasi ASL o altro soggetto (Regione, Agenzia, . . . ) che integra dati amministrativi sanitari

**Tempi** Il software verrà rilasciato entro la fine del progetto (gennaio 2014)



# Limiti e rimedi

**Record linkage deterministico** Il record linkage implementato da TheMatrix è deterministico, quindi i suoi dataset riproducono i problemi di qualità delle sorgenti originali

**Algoritmi e validazione** Per individuare le variabili mancanti il ricercatore implementa in TheMatrix degli algoritmi, quindi deve validarne la qualità o verificare che ci siano degli studi di validazione affidabili



## Un monito

Di fronte alle grandi potenzialità del record linkage ci si ritrova in un dilemma



## Un monito

Crederne nei dati senza porsi  
troppe domande. . .



## Un monito

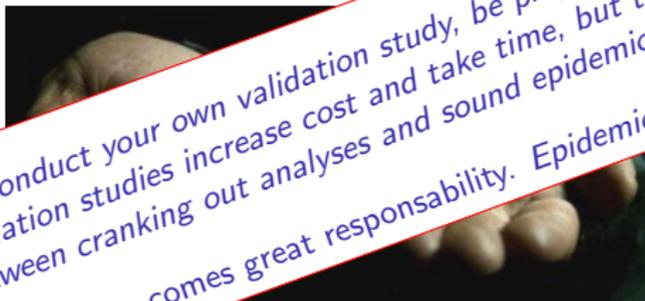


... oppure mettere alla prova i  
dati per verificare quanto sono  
lontani dalla realtà

# Un monito



# Un monito



If you did not conduct your own validation study, be prepared to cite others who did. Validation studies increase cost and take time, but they may be the difference between cranking out analyses and sound epidemiologic research.

Hernán M. With great data comes great responsibility. Epidemiology 2011 May.

# Caratteristiche

**Portabilità** Applicazione Java, accede via driver standard (JDBC) indifferentemente a MySQL, Oracle, Microsoft SQLServer, SAS

**Flessibilità** Può leggere anche dati in formato csv, con qualsiasi tracciato record

**Script** La generazione di ogni dataset è governata da uno script sviluppato ad hoc nel linguaggio di programmazione di TheMatrix

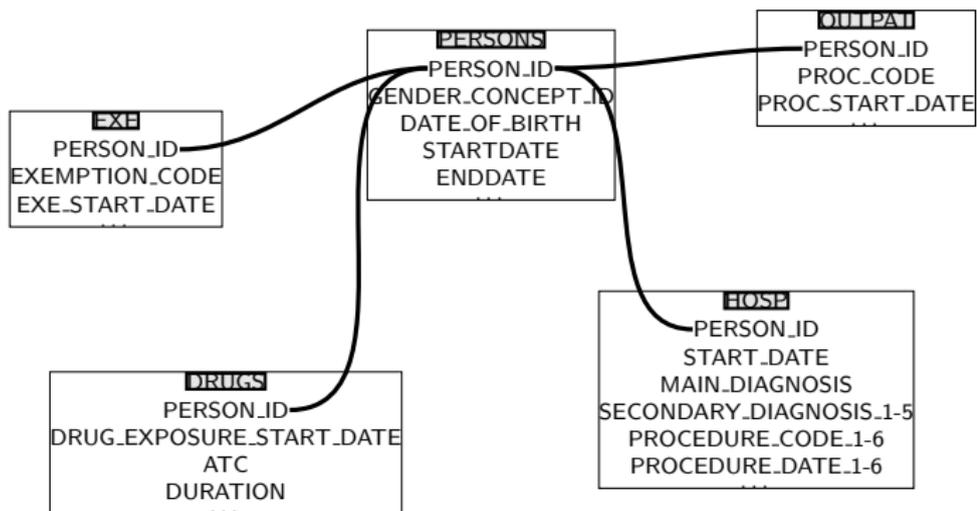
**Programmazione script** Tramite interfaccia grafica

**Poca informatica** L'intervento di una professionalità propriamente informatica può essere limitato al momento dell'installazione

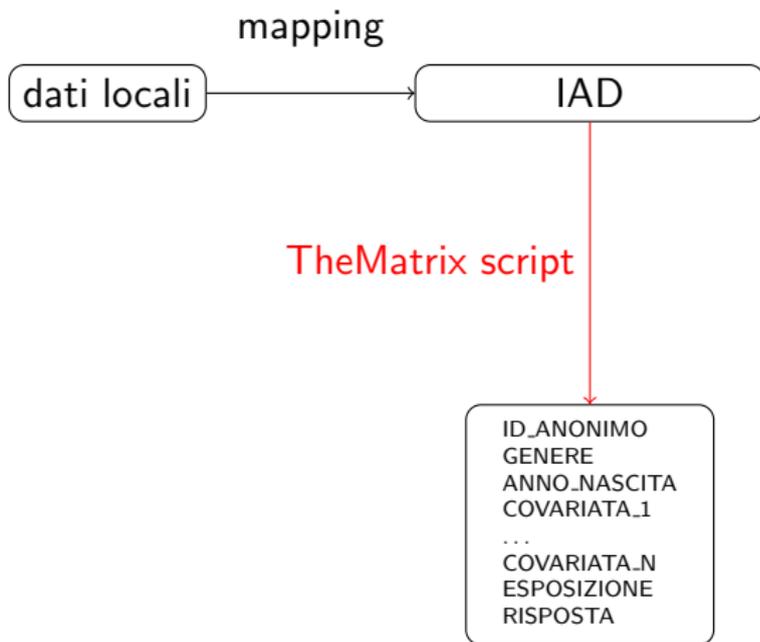
**Output** Il formato dell'output è il semplice csv



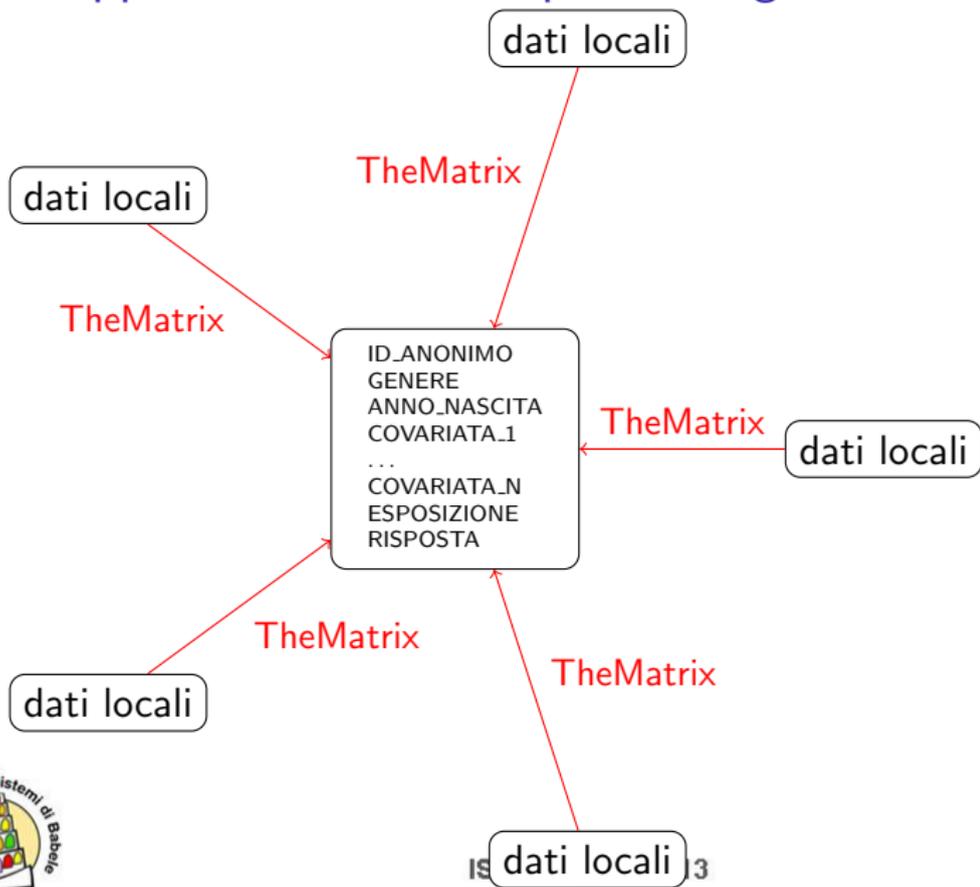
# Database virtuale: IAD



# Architettura



# Applicazione a studi epidemiologici multicentrici



# Indice

Introduzione

Caratteristiche e logica di TheMatrix

Obbiettivi e funzionalità

Linguaggio di programmazione e interfaccia grafica

Requisiti e modalità di installazione

Conclusioni



# Espressività

**Analogie** Simile al linguaggio di data management Stata, SAS, e altri

**SQL elementare** Aggrega, filtra, integra...

**Più complesse** Filtra su un prodotto cartesiano senza calcolarlo tutto

**Manipolazione delle colonne** Aritmetica, operazioni su date, ...



## Frammento di script

```
NumberFirst_DIAB_DRUGS (AggregateModule)
inputs
custom = YearFirst_DIAB_DRUGS
parameters
isInputSorted = false
groupBy = [ PERSON_ID,YearFirst_DIAB_DRUGS ]
functions = [ COUNT( PERSON_ID ), MIN(DRUG_EXPOSURE_START_DATE) ]
results = [{NumberFirst_DIAB_DRUGS:int};{DateFirst_DIAB_DRUGS:date}]
end
FilterTwoFirst_DIAB_DRUGS (FilterModule)
inputs
custom = NumberFirst_DIAB_DRUGS
parameters
conditions = [{NumberFirst_DIAB_DRUGS>1}]
end
```



# Interfaccia grafica

<http://www.ars.toscana.it/thematrix.html>



# Indice

Introduzione

**Caratteristiche e logica di TheMatrix**

Obbiettivi e funzionalità

Linguaggio di programmazione e interfaccia grafica

Requisiti e modalità di installazione

Conclusioni



# Requisiti

**Sistema operativo** Supportati Linux e Windows

**Java** Sun Java 1.5 o superiore

**Spazio disco** Al massimo: circa 3 volte il database originario, in generale dipende dallo script

**Accesso ai dati** Due alternative

**Virtuale** Accesso via JDBC in lettura (supportati MySQL, Oracle, Microsoft SQLServer, SAS)

**Fisico** Accesso a una copia dei dati in formato csv nello schema IAD



## Frammento di file di mapping

```
<iadMapping>                                <!-- mapping of PERSON -->
  <dataset name="PERSON">
    <joinName>PERSON</joinName>
    <joinClause>FLUSSI.ANAGRAFE</joinClause>
    <simpleMapping name="BIRTH_LOCATION_CONCEPT_ID">
      <sourceTable>FLUSSI.ANAGRAFE</sourceTable>
      <sourceAttribute>COMNASC</sourceAttribute>
    </simpleMapping>
    <simpleMapping name="DATE_OF_BIRTH">
      <sourceTable>FLUSSI.ANAGRAFE</sourceTable>
      <sourceAttribute>DATANASC</sourceAttribute>
    </simpleMapping>
```

...



# Indice

Introduzione

Caratteristiche e logica di TheMatrix

Conclusioni



## Riassumendo

**Software open source** TheMatrix è un software open source che verrà rilasciato entro il gennaio 2014

**Progetto MATRICE** TheMatrix è un prodotto del progetto MATRICE dell'Agenas

**Database virtuale** TheMatrix crea un database virtuale basato sui dati sanitari amministrativi ed estendibile tramite un linguaggio di scripting

**Facilmente programmabile** Gli script di TheMatrix si possono programmare con l'aiuto di un'interfaccia grafica

**Installazione** L'installazione si può fare su molti sistemi, eventualmente con il supporto di un informatico

**Validazione** Attenzione a usare algoritmi validati! E a interpretare i risultati di conseguenza



Grazie per l'attenzione!

