DELPHI

Data e-Platform to Leverage Multilevel Personal Health Information

Yannis Katsis
CSE, UCSD

Kevin Patrick, Natasha Balac, Chaitan Baru, Jessica Block, Ted Chan, Sanjoy Dasgupta, Claudiu Farcas, William Griswold, Jeannie Huang, Yannis Katsis, Massimiliano Menarini, Lucila Ohno-Machado, Yannis Papakonstantinou, Fredric Raab, Besa Smith
DELPHI Overview
Multiple sources of health data

- Medical Records
- Personal Health Data (weigh-ins, run info, ...)
- Environmental Data (pollution, noise, greenspace, ...)
- Genomic Data
- Microbiome Data
- Public Health Data
Multiple sources of health data

...of different types

<table>
<thead>
<tr>
<th>Behavioral &amp; Social Data</th>
<th>Physical activity</th>
<th>Sleep</th>
<th>Social Networks</th>
<th>Stress</th>
<th>Diet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Medical Data</td>
<td>Medical Records</td>
<td>Genomic</td>
<td>Pharmaceutical</td>
<td>BMI</td>
<td></td>
</tr>
<tr>
<td>Environmental Data</td>
<td>Food &amp; Grocery</td>
<td>Pollutants</td>
<td>Transportation</td>
<td>Crime &amp; Incivilities</td>
<td></td>
</tr>
</tbody>
</table>
Providing health care & population health requires reasoning across these layers

- **Behavioral & Social Data**
  - Physical activity
  - Sleep
  - Social Networks
- **Personal Medical Data**
  - Medical Records
  - Genomic
  - Pharmaceutical
  - BMI
- **Environmental Data**
  - Food & Grocery
  - Pollutants
  - Transportation
  - Crime & Incivilities

- **Diabetes Management**
- **Asthma Care**

**Obesity tracking for public health**
Today: Most health data ignored

Reasons:

- Data are collected and maintained by different agencies
  - Making it hard to find and access them
- Data have different data types
  - Making it hard to combine them
DELPHI: The Goal

- Integrate heterogeneous data into a “single” uniform database
  - By taking into account the geospatial context
- Implement an analytics and visualization layer on top
- Open data and analytics to 3rd-party developers of apps & services

Enable personalized population health through the creation of a “Whole Health Information Platform” that takes into account everything from the genome to the exposome – essentially all health-relevant data

Partners
DELPHI System Architecture

**Sources**
- Medical Records*
- Personal Sensor Data*
- Location Data*
- Smartphone
- Qualcomm Life
- BEACON Health Info Exchange
- Genomic Data
- Environmental Data
- 3rd party web-services

**DELPHI**
- WHOLE HEALTH INFORMATION MODEL (WHIM)
- ANALYTICS LAYER: Allows developers to run common analytics efficiently
- VISUALIZATION LAYER: Allows developers to create common visualizations efficiently

**Applications**
- Patient/Parent App
- Medical Personnel Individual’s Dashboard
- Medical Personnel Population Dashboard
- Other New Applications
  - Feedback, alert & advice
  - Integrated view of patient + feedback
  - Population statistics & analytics

**Goal:** Create ecosystem for developers to create the next generation of health applications

**WHAPI:** Allows developers to access all integrated health data and write apps that use them.
Use Case: Asthma

Goal: Improve asthma care by combining environmental, activity, and hospitalization data

- **Environmental Data**
  - Air quality, County health
- **Personal (Sensor) Data**
  - Activities, Peak Flow, Self Report
- **Medical Records**
  - Hospitalization, Age, Flu shot

**DELPHI**

- **Patient/Parent App**
  - Probability of danger (machine learning algorithms) + contact to doctor
- **Medical Personnel Individual’s Dashboard**
  - Integrated view of patient + feedback
- **Medical Personnel Population Dashboard**
  - Population statistics

**PALMS**

- Activities, Peak Flow, Self Report
Asthma Application
Use Case: Behavioral Health

Goal: Find indicators correlated with behavioral health outcomes

- Anxiety Disorders
- Mood Disorders
- Alcohol-related Disorders

Physical & Social Environment:
- Diet
- Smoking
- Stress
- Sleep
- Activity
- Near Open Space
- Near busy Roadways
- Pesticides
- Access to Healthy Food
- Drinking Water Violations
- Traffic
- Air Pollution
- Crime Rate
- Transportation to Work

Health Systems:
- Uninsured
- Immunization Status
- Diabetic Screening
- Uninsured
- Highschool Graduation
- Children in Poverty
- Marital Status
- Income Inequality

Social & Economic Conditions:
- Housing Problem
- Unemployment Rate
- Overcrowding
- Shelter Services
- Financial Aid

Public Policies:
- Highschool Graduation
- Children in Poverty
- Marital Status
- Income Inequality

Individual Behaviors:
- Activity
- Sleep
- Stress
- Smoking
- Diet

...
DELPHI DB: An Overview

- A PostgreSQL database
- With two schemas
  - cogs121_16_integrated: All indicators integrated into a uniform format
  - cogs121_16_raw: Raw indicator tables
Integrated DB: An Overview

- Contains ~3,000 indicators/variables
- From 4 datasets
  - 2013 Market Potential Data: Consumer buying patterns
  - 2012 HHSA Demographics: Demographics
  - 2012 HHSA Behavioral Health Outcomes: Hospitalizations & emergency department discharges for behavioral health conditions
  - 2012 SANDAG Healthy Communities Assessment Atlas: Information on physical & built environment (e.g., access to parks, libraries, etc.)
- With values for 41 Subregional Areas (SRAs) in San Diego County
  - An indicator may not contain values for some SRAs
- Of different types
  - Default: Absolute number of people with some characteristic
  - “Rate”: Number of people per 100,000 population
  - “Percent”: Number of people per 100 population
Integrated DB: The Schema

**datasets**
- PK: dataset_id
- dataset_name

**indicators**
- PK: indicator_id
- FK: dataset_ref
- indicator_name
- data_type

**sras_to_indicators**
- FK: sra_ref
- FK: indicator_ref
- value_string
- value_integer
- value_double

**sras**
- PK: sra_id
- sra_name
Raw DB: An Overview

- ~300 tables
- Each containing multiple indicators
- Contains data not contained in the integrated schema:
  - San Diego Air Pollution Control District (SDAPCD): Air quality data
  - Environmental Protection Agency (EPA): Air quality data
  - Health & Human Services (HHSA): Additional years of behavioral health data and health outcome data about other conditions (chronic diseases, injuries, etc)
  - Zillow: Housing market data
  - San Diego Association of Governments (SANDAG): Environmental point data
  - Automated Regional Justice Information System (ARJIS): Crime data
- Contains data with finer spatial granularity:
  - zip codes, points, polygons, ..
- Contains metadata
  - Table ‘metadata_tables’ contains description of each table
  - Table ‘metadata_sources’ contains the data sources