Integration and analysis of heterogeneous big data for precision medicine and suggested treatments for different types of patients.

iASiS: Big Data to Support Precision Medicine and Public Health Policy

Maria-Esther Vidal

Leibniz Information Centre For Science and Technology University Library (TIB) and L3S Institute @ Leibniz University of Hannover (LUH), Germany

Coordinator: NCRS Demokritos, Greece (George Paliouras)
## Motivation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>38%</td>
</tr>
<tr>
<td>Asthma</td>
<td>40%</td>
</tr>
<tr>
<td>Cardiac Arrhythmias</td>
<td>40%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>43%</td>
</tr>
<tr>
<td>Migraine</td>
<td>48%</td>
</tr>
<tr>
<td>Arthritis</td>
<td>50%</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>52%</td>
</tr>
<tr>
<td>Alzheimer’s</td>
<td>70%</td>
</tr>
<tr>
<td>Cancer</td>
<td>75%</td>
</tr>
</tbody>
</table>

iASiS Vision:  
**Turn** clinical, pharmacogenomics, and other **Big Data** into actionable **knowledge** for personalized medicine and health policy-making

Objectives:  
• **Integrate** automated **unstructured** and **structured** data analysis, **image** analysis, and **sequence** analysis into a **Big Data framework**

• **Develop** a **framework** for supporting **personalised** diagnosis and treatments
Pilot 1: Lung Cancer

Motivation:

- **Lung cancer** among the most
  - *Common* and *deadly* diseases
  - *Costly* cancers
- **Lung cancer** is a *heterogeneous* disease. Characteristics *differ* among
  - Patients
  - Tumor regions

iASiS will enable:

- **Discovery** of correlations among tumor spread, prognosis, response to treatment
- **Unraveling** molecular mechanisms that *predict response* to different tumor types (signatures)
Motivation:

- Approximately, 10% of people over 65 suffer from Alzheimer’s
- **Heterogeneity** of the symptoms impedes accurate diagnosis and treatments

iASiS will enable:

- **Discovery** of **patterns associated** with prognosis, outcomes, and response to treatments
- **Association** of **medical and lifestyle advice** with Alzheimer’s risk and stages of severity
The iASiS Pipeline

Input
- EHR Text Analysis
- Genomic Analysis
- Image Analysis
- Open Data Analysis

Knowledge Extraction
- Annotated Datasets

Knowledge Graph Creation
- Semantic Enrichment
- Knowledge Curation & Integration
- Interlinking

Knowledge Management & Discovery
- Federated Query Processing
- Pattern Discovery
- Link Prediction
- Relatedness Discovery

Output
- Actionable Knowledge
  - P53
  - CALGB 9633
  - KRAS
  - Lung Tumor Signature
  - Drug-Target Interactions
  - Long Survival Patients

Enforcing Data Access Control and Data Privacy Policies
How the Community can contribute?

Datasets:

Pharmacogenomics and Clinical datasets:
- Information about Lung cancer or Alzheimer’s patients
- Clinical record data, images related to clinical records, and genomic data

Data Analytics Tools:

Data analytics tools able to:
- Discovery communities of similar entities
- Predicting associations in the iASiS KG
- Determining relatedness between entities in iASiS KG
How iASiS will contribute to the community?

iASiS Knowledge Graph (KG)

**RDF** knowledge graph integrating:
- Information about **Lung cancer** or **Alzheimer’s** patients
- **Pharmacogenomics** data
- **Interactions between** drugs, drugs and proteins, and drugs and side effects.

Data Analytics Tools:

- **Data analytics** tools able to:
  - Predict interactions between entities in the iASiS KG
  - **Uncover unknown** patterns in patients with **Lung cancer** or **Alzheimer’s disease**.
Available Tools

**GADES**, a *semantic similarity measure* to determine relatedness between entities in a knowledge graph

https://github.com/RDF-Molecules/GADES

**semEPNode**, a community detection tool for knowledge graph partitioning


**A demo reporting** the behavior of state-of-the-art approaches for predicting interactions between drugs and targets

https://project-demo-bayer.herokuapp.com/indexpage/index.html
Partners Involved

THANK YOU!