



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



<http://project-iasis.eu>



@Project_IASIS

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

Anastasia Krithara

**Inst. of Informatics & Telecommunications
NCSR “Demokritos”, Greece**



SC1-PM-18-2016: Big Data supporting Public Health policies

iASiS Basic Facts

- Title: Integration and analysis of **heterogeneous big data** for **precision medicine** and suggested **treatments** for different types of patients
- Topic: H2020-SC1-PM-18-2016 - Big Data supporting Public Health policies
- Contract No.: 727658
- Budget: € 4.3M



Motivation

- Epidemiological data analysis is not sufficient for public health policies in the era of personalized/precision medicine
- We also need explanations, e.g. why a treatment ought to work better for one type of patient than another
- Therefore, we need to combine **breadth** (across a population) with **depth** (e.g. personal genome) in the analysis
- Big data analysis can address both breadth and depth, under the appropriate framework. **That's iASiS!**



Vision and Objectives

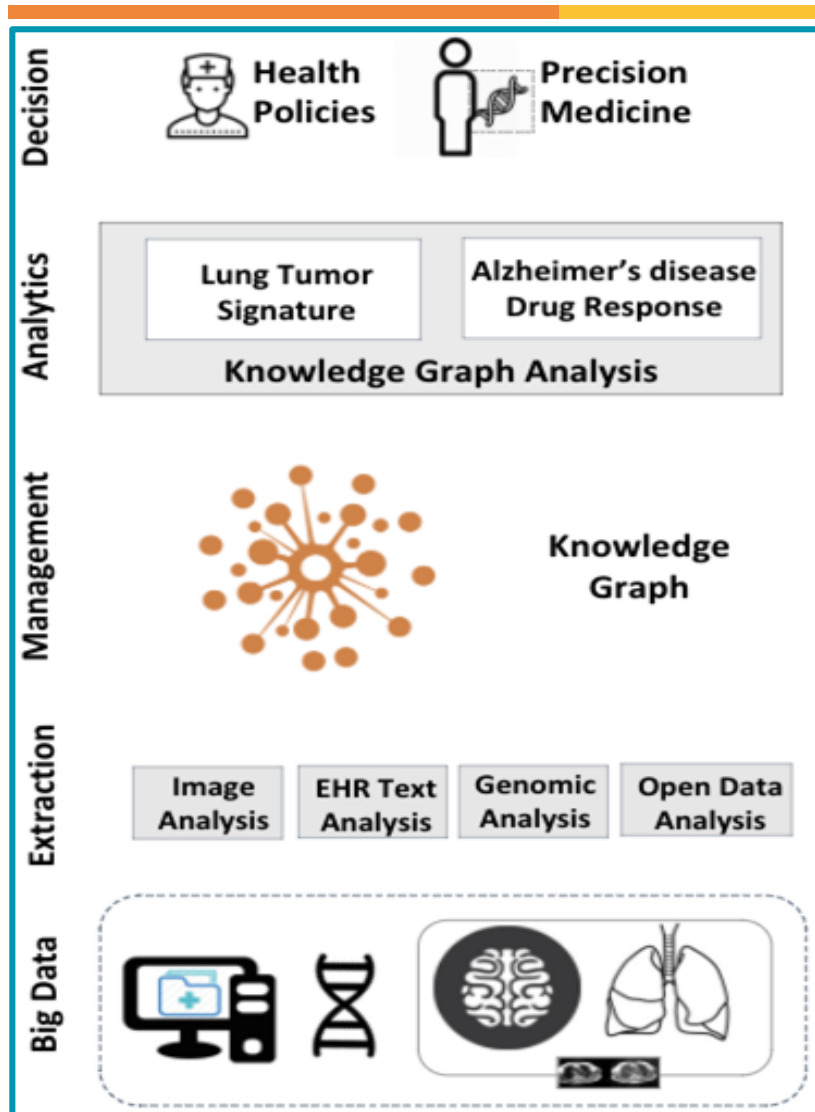
iASiS Vision:

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

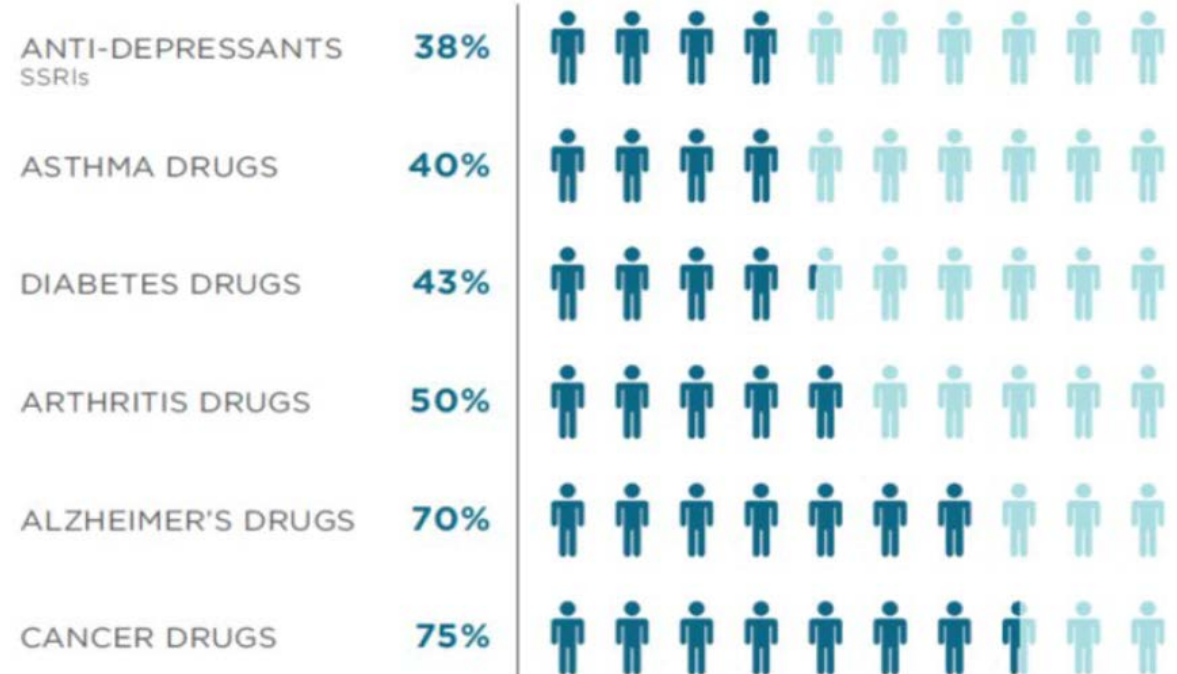
iASiS Objectives:

- Integrate automated **unstructured** and **structured** data analysis, **image** analysis, and **sequence** analysis into a **Big Data** framework
- Use the iASiS framework to support **personalized diagnosis and treatment**

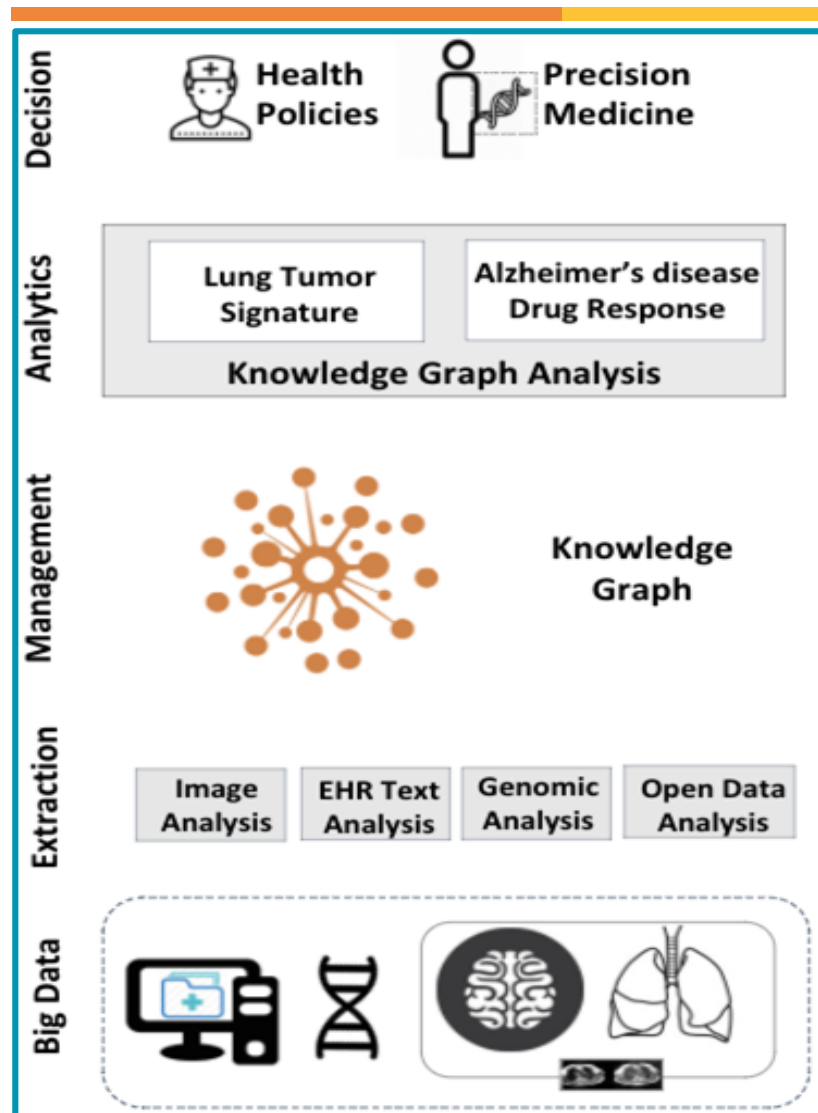
The iASiS Framework



- iASiS focuses on **two use cases**:
 - Lung cancer
 - Alzheimer's disease
- General-purpose drugs are often ineffective

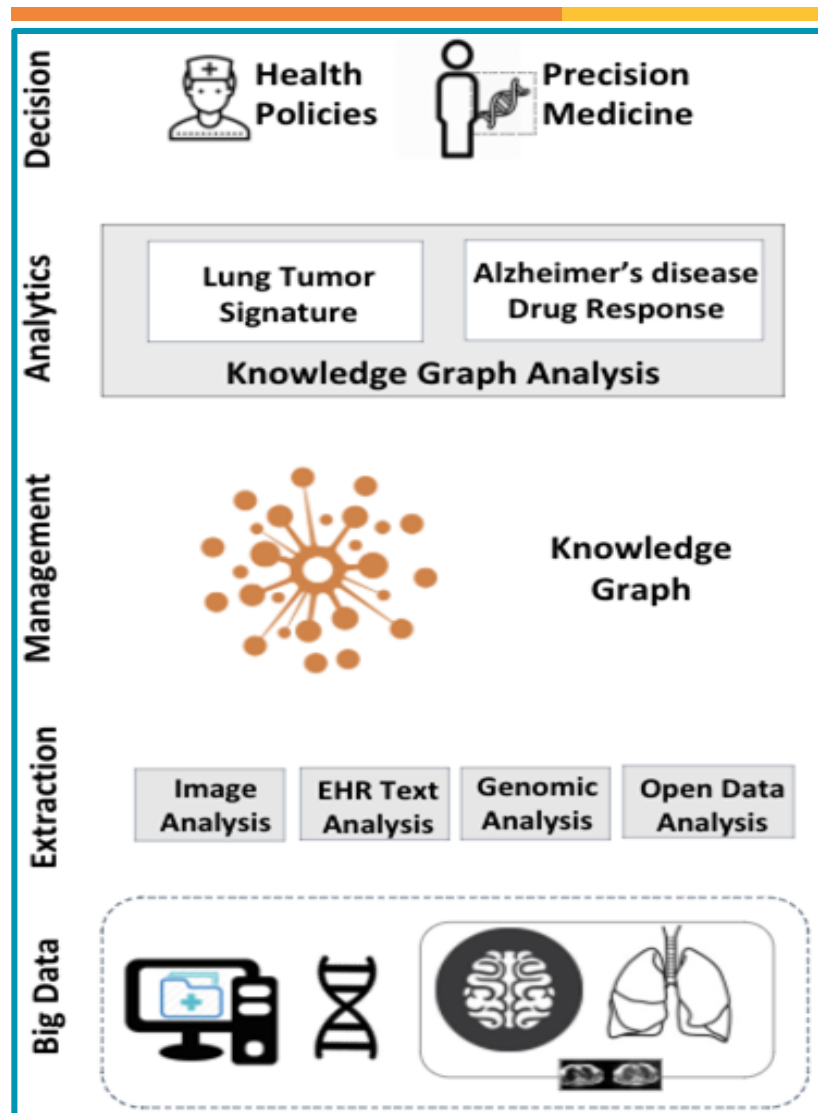


The iASiS Framework



- iASiS **analyses**:
 - EHRs (English & Spanish)
 - MRI & PET/CT images
 - Genomic data (e.g. liquid biopsy samples)
 - Related bibliography (e.g. PubMed)
 - Biomedical databases (e.g. DrugBank)
 - Biomedical ontologies (e.g. GO, UMLS)

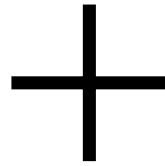
The iASiS Framework



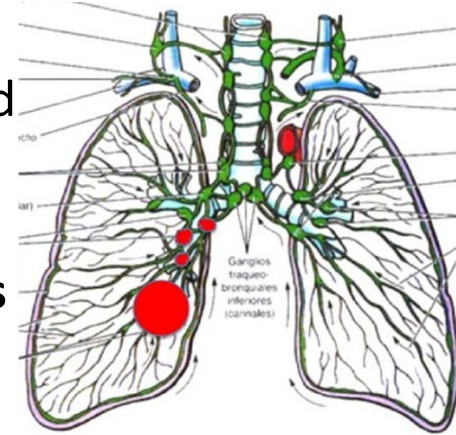
- Extracted knowledge is fused in the iASiS **knowledge graph**
 - Unified semantic schema
 - Linked data
 - Machine-processable knowledge
- iASiS **end-users can:**
 - Perform natural language questions
 - Receive answers along with justifications
 - Identify patterns in patient populations
 - Make informed decisions
- All steps of data management and analytics enforce **privacy** and **access** control

Lung Cancer Pilot Data

- EHRs in Spanish
- PET/CT Images
- Genomic Data/Liquid Biopsy Samples



- Pharmacological knowledge extracted from **publicly available datasets**
- **Biomedical ontologies and taxonomies**
 - terminology standardization
 - semantically describing the EHRs



Hospital Universitario
Puerta de Hierro
Majadahonda



EUROPEAN
GENOME-PHENOME
ARCHIVE



NATIONAL CANCER INSTITUTE
Genomic Data Commons



GENEONTOLOGY
Unifying Biology



Unified Medical
Language System®

Lung Cancer Use case

Improvement of treatment schemes for lung cancer

Input

Patient data:

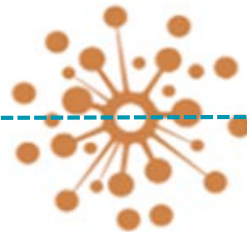
- Clinical notes
- clinical reports
- age at diagnosis
- stage and type of tumour
- gender
- ...

Output

1. Patient categorization in long survival group based on discovered correlations:

- females vs males survival
- risk factors, like high blood pressure, smoking habits and COPD

Optimal personalised treatment schemes



Analysis

- Analysis of patient clinical records (EHR, images, liquid biopsy data)
- Analysis of open data (related publications, ontologies, structured databases)

Aim: improve treatment schemes, based on long surviving patients

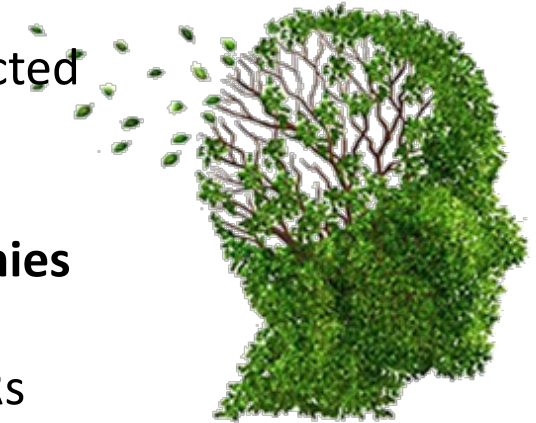


online

offline

Alzheimer's Disease Pilot Data

- EHRs in English
 - MRI Brain Images
 - Genomic Data
- +
- Pharmacological knowledge extracted from **publicly available datasets**
 - **Biomedical ontologies and taxonomies**
 - terminology standardization
 - semantically describing the EHRs



CRIS
NETWORK



Alzheimer's use case

Optimal choice of drug in Alzheimer's disease

Input

Patient's data:

- Disease stage (MMSE)
- Allelic status
- Cognitive enhancing drugs
- Age, Sex
- Comorbidities
- Lifestyle
- Family history

Output

1. Positive response estimates for each available drug: Donepezil, Galantamine, Rivastigmine etc.
2. Supporting data:
 - Publications that support the results
 - Statistics for similar patients

Decision on drug prescription



Analysis

- Analysis of patient clinical records (EHR, images, genotype)
- Analysis of open data (related publications, ontologies, structured databases)

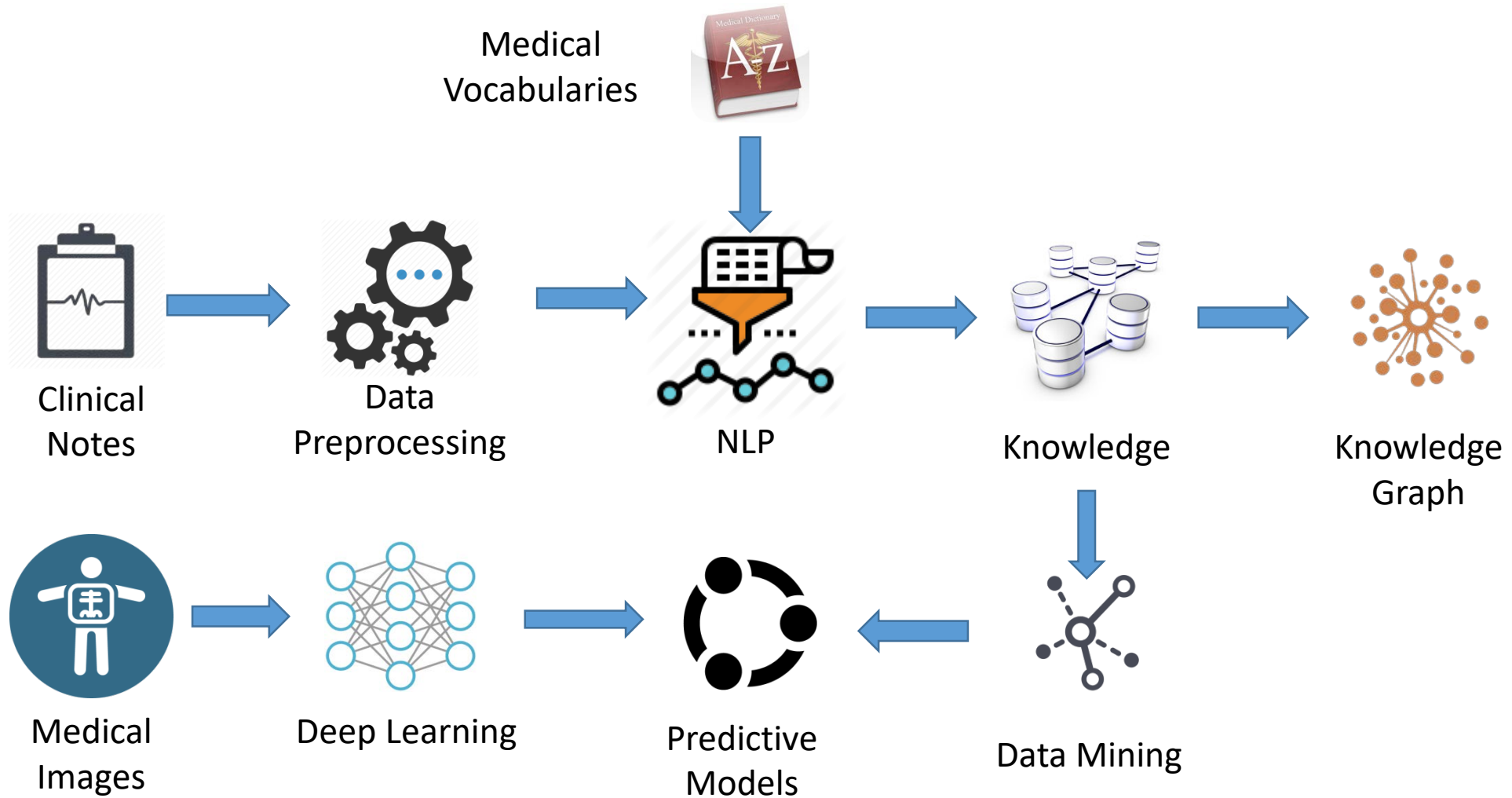
Which of the available drugs is most suitable for a particular patient?



online

offline

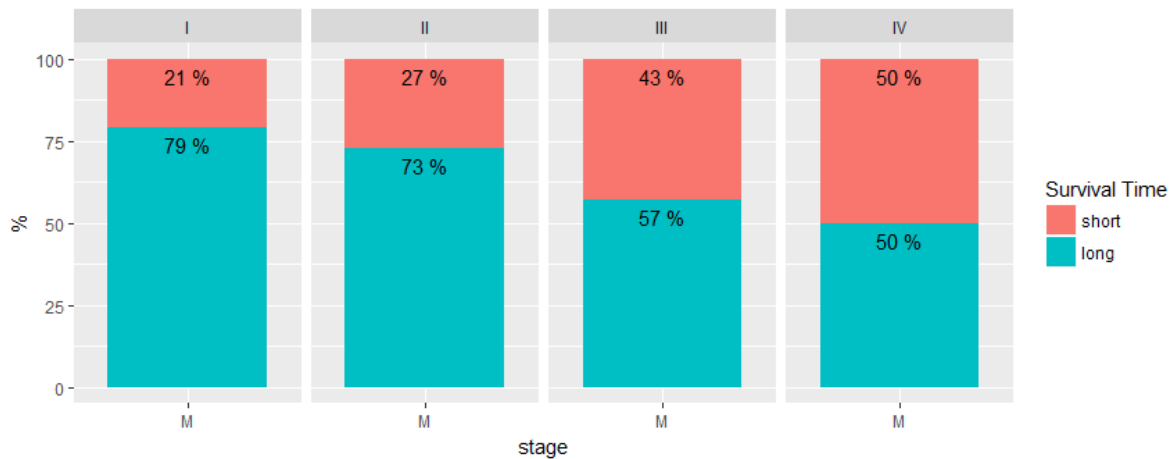
Clinical Data Analysis



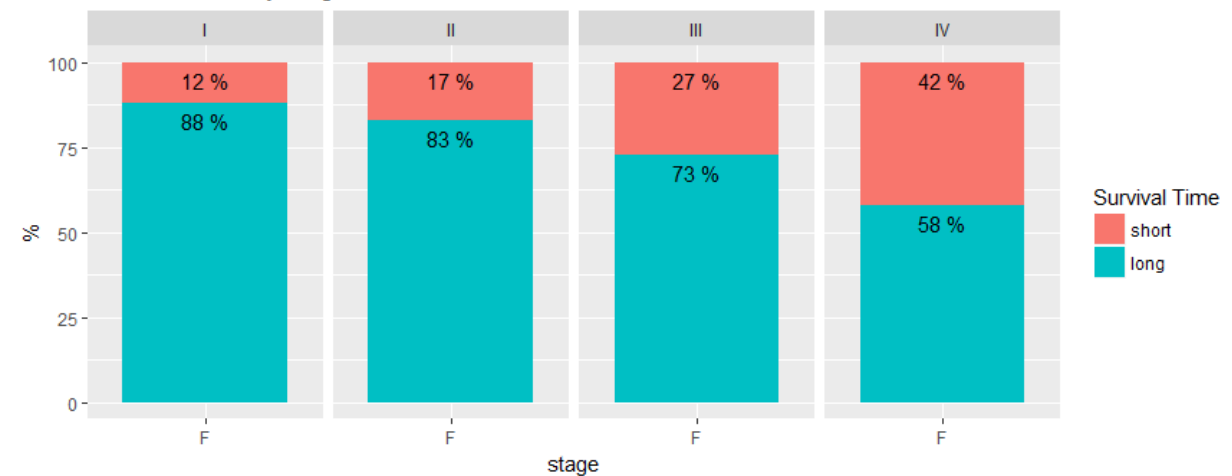
Preliminary Results – Lung Cancer

- Analysis of more than 170,000 clinical notes and 7,000 clinical reports from 706 patients
- Correlation was observed between presence of risk factors, such as dyslipidemia, high blood pressure, smoking habit and COPD, and significant decrease in survival.
- Survival in females was significantly higher than in males, despite the smoking habit, stage and comorbidities

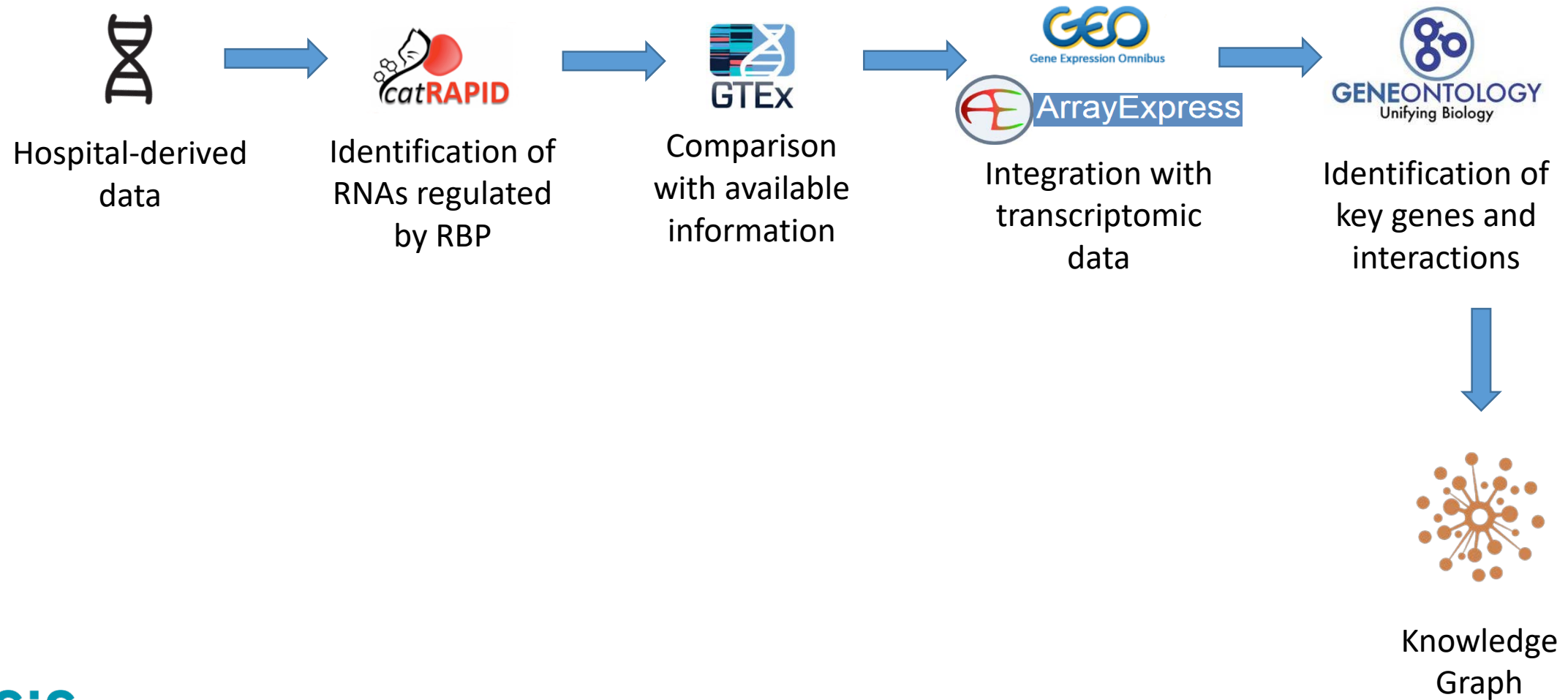
Survival Time by stage in Males



Survival Time by stage in Females



Genomic Data Analysis



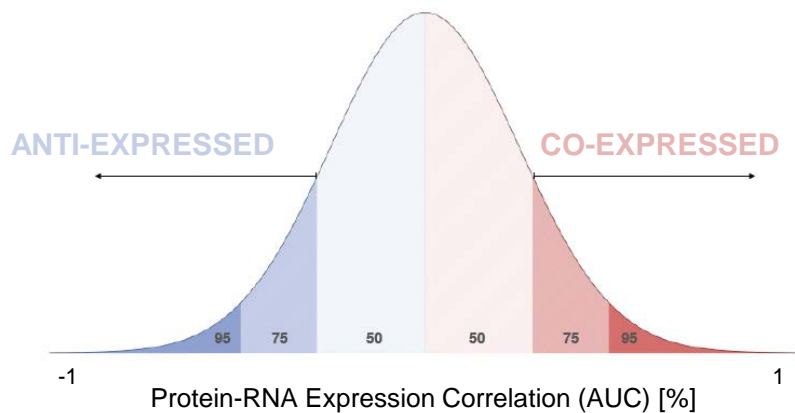
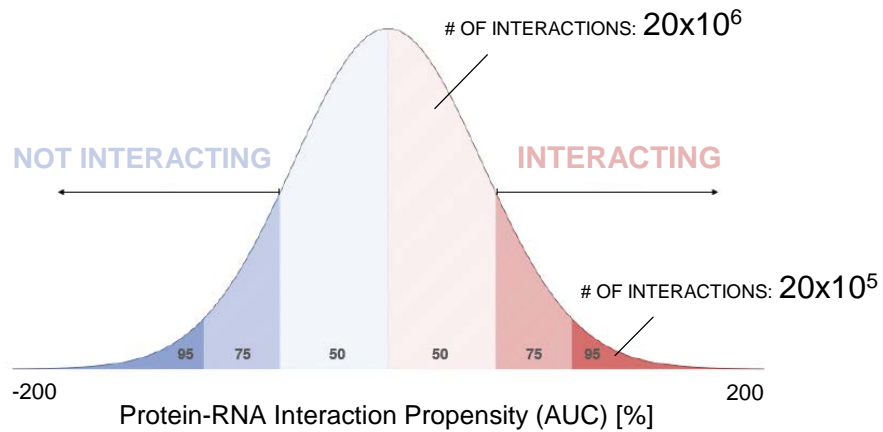
Genomic Data Analysis



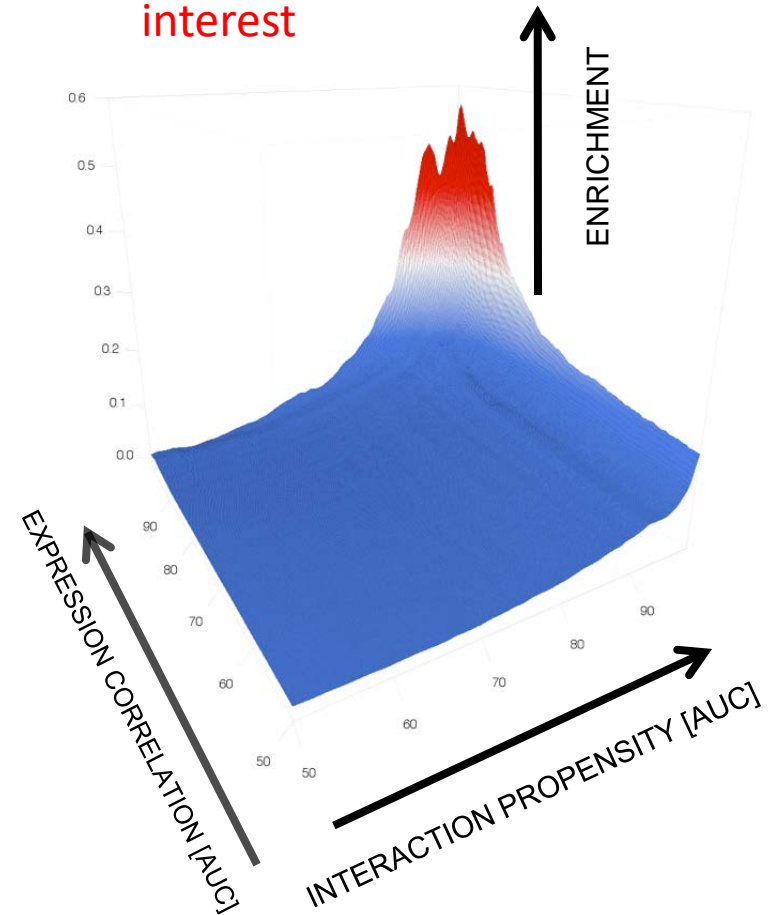
catRAPID



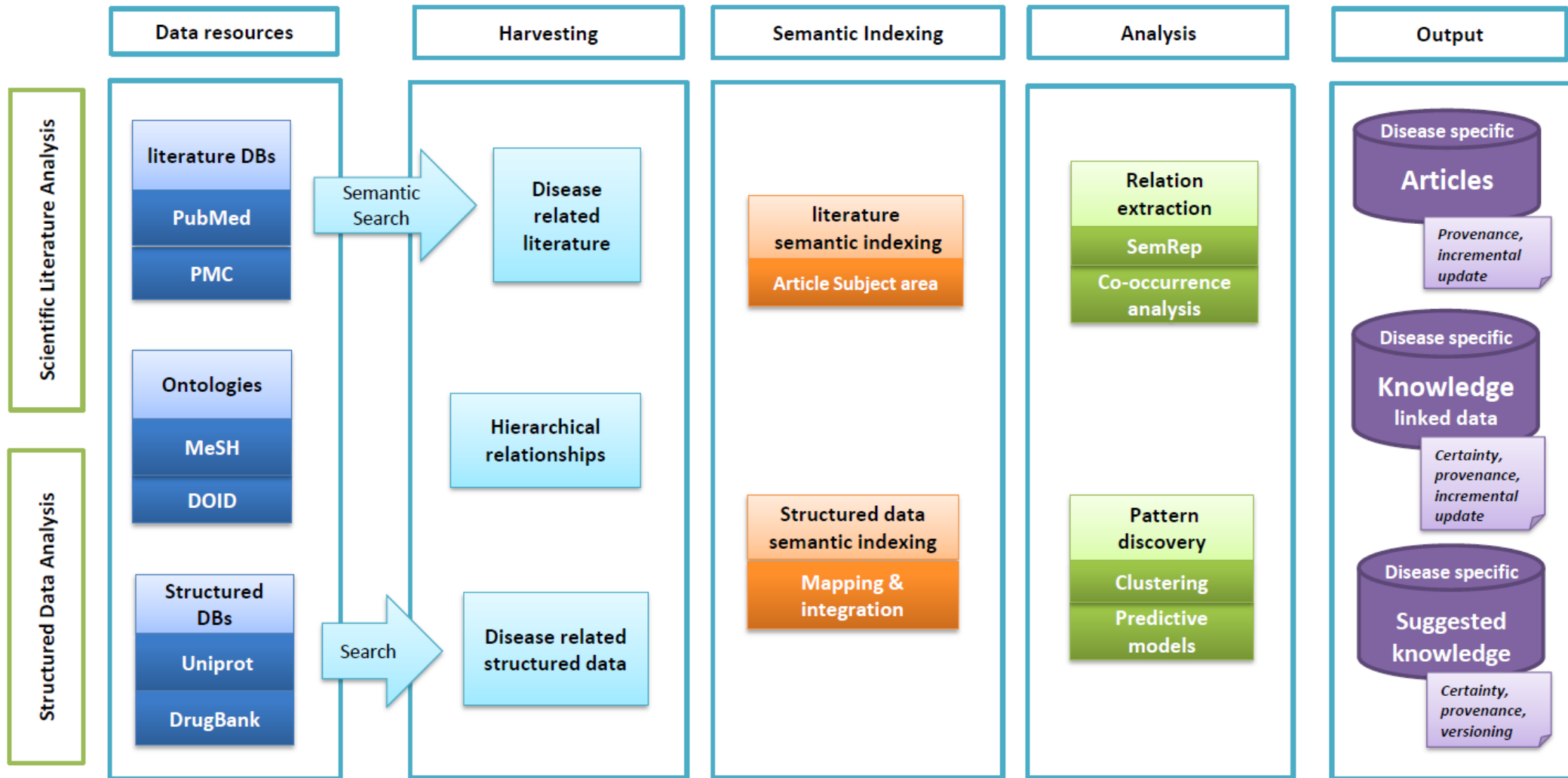
RNAseq



The correlation between co-expressed protein-RNA pairs (RNAseq) and their interactions (catRAPID) identifies cases of interest



Open Data Analysis



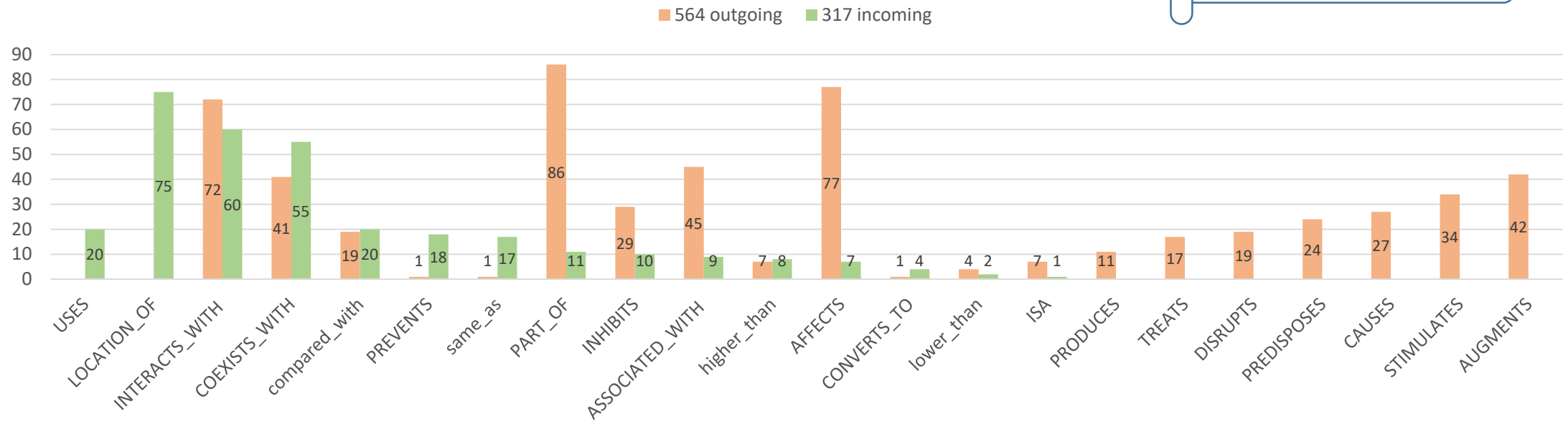
Preliminary Results –Alzheimer’s

Identification of alleles of risk with related treatments according to the current bibliography

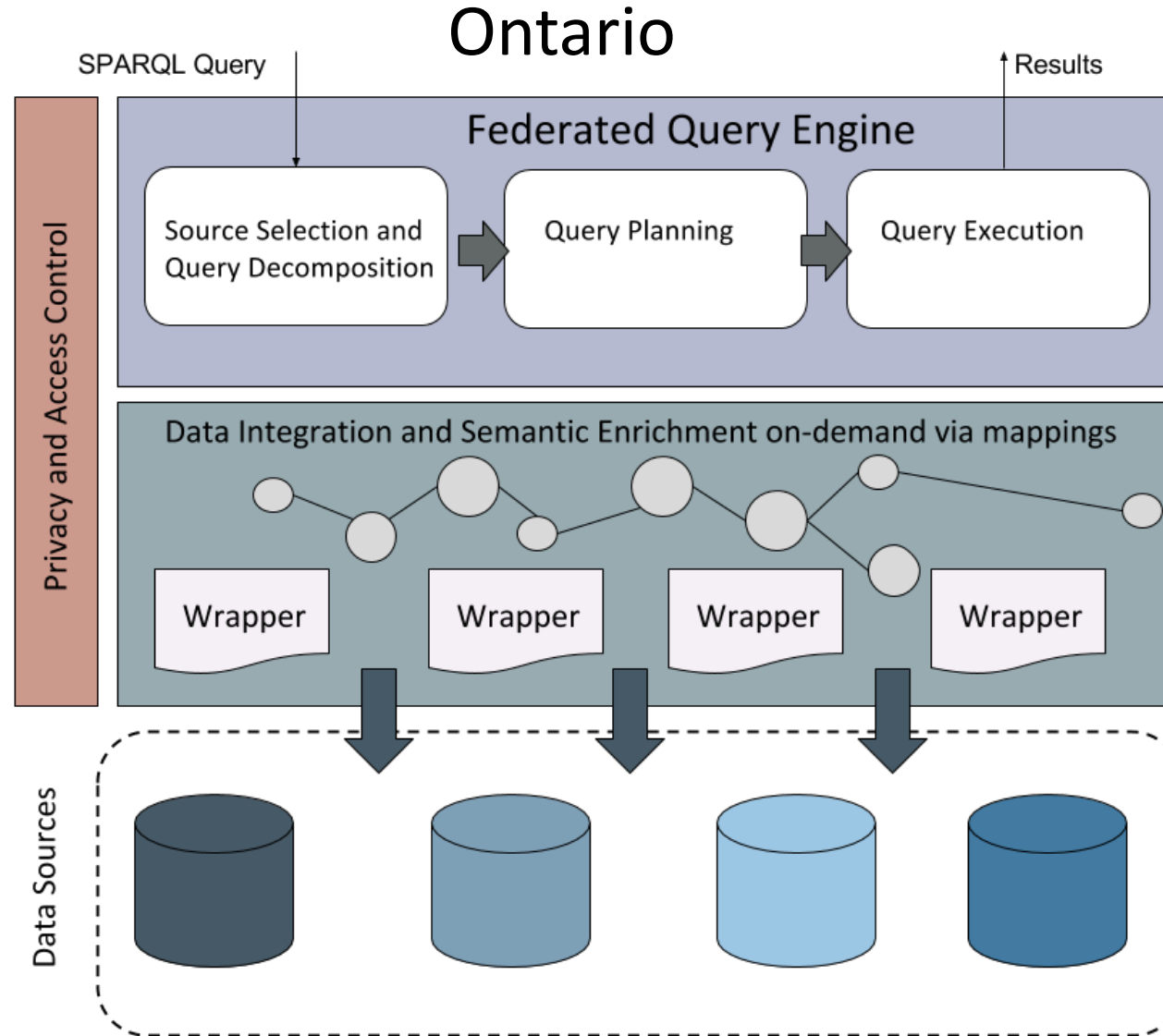
- APOE4 in AD relevant literature: **2,201 articles** have APOE4 as **Topic** / APOE4 occurs in **1,908 articles**
- Knowledge about APOE4: **883 distinct relations** of 22 types between APOE4 and 561 concepts
 - 317 incoming relations (APOE4 as object, e.g. “Therapeutic procedure **USES** APOE4”)
 - 564 outgoing relations (APOE4 as subject, e.g. “APOE4 **CAUSES** Asthenia”)

In $\sim 2 \times 10^3$ out of $\sim 100 \times 10^3$ articles “relevant to AD”

Extracted relations for APOE4



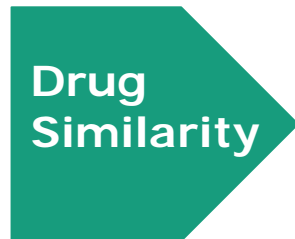
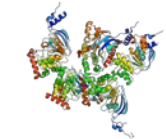
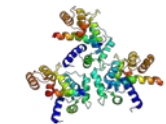
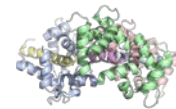
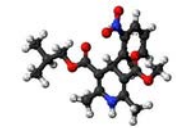
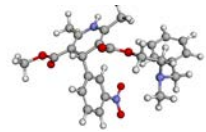
Data Management



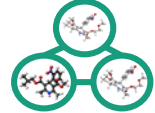
Knowledge Graph

High-Level Analysis

Computing Similarity



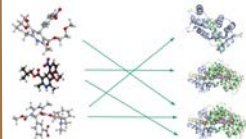
ChEBI Ontology



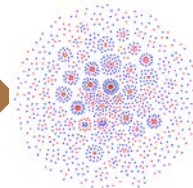
Graph Partitioning



Configuration Parameters

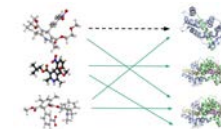
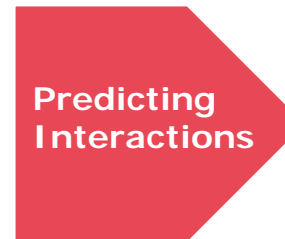


Knowledge Graph



Clusters of similar Drugs, similar Targets, and their interactions

Prediction Principle

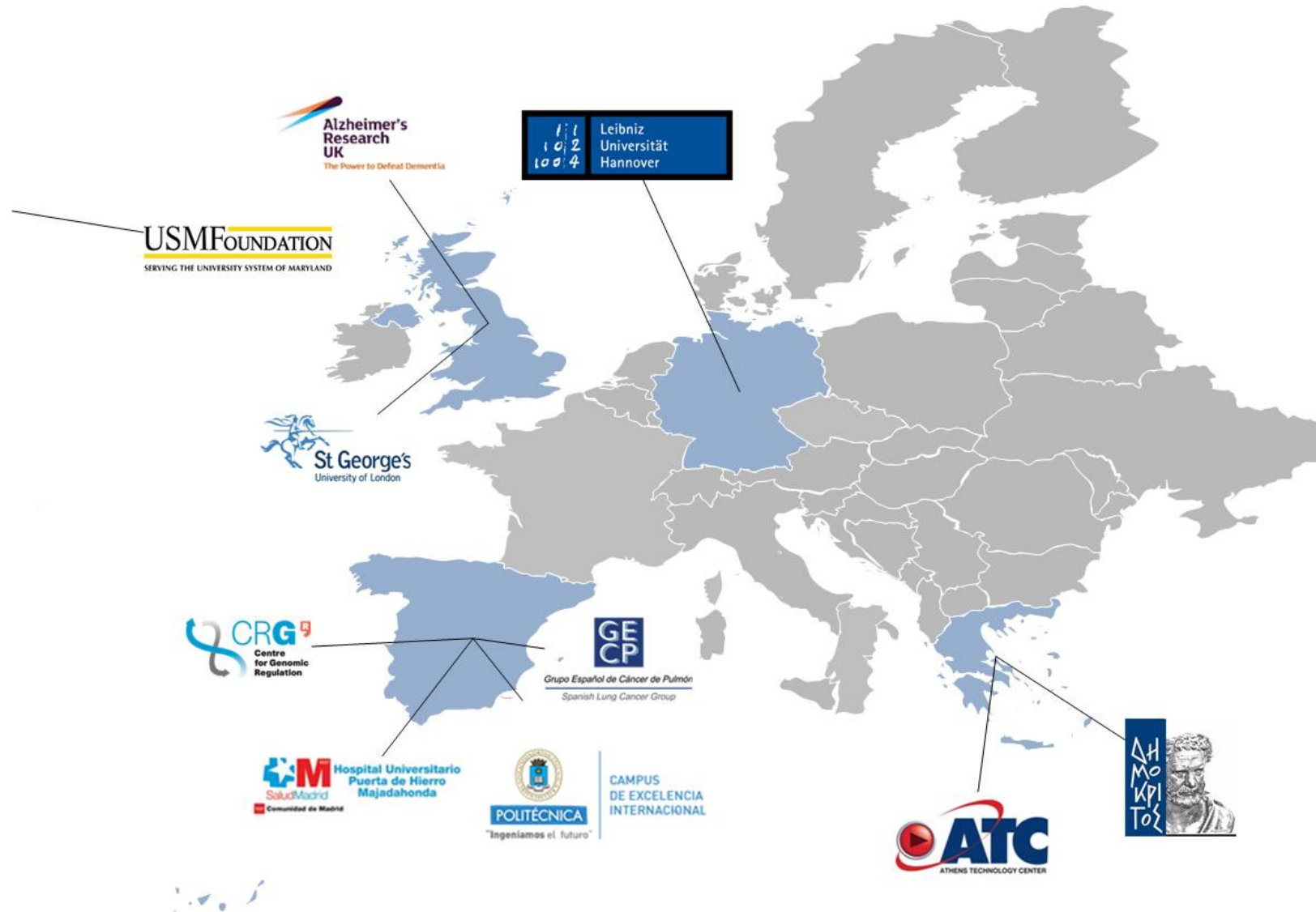


Discovered Links and Patterns

Beyond Data Analysis

- iASiS handles **sensitive patient data** from hospitals: EHRs, MRI and PET/CT images, blood and liquid biopsy samples
- Ethics Committee led by **external advisor** to oversee the adherence to rules, regulations and patient consent per data source.
- Data management plan using **FAIR principles** and corresponding tools.
- Data **access control**, including anonymization, hardware and software protection, regulated access.

iASiS Partners



Upcoming event

Symposium on "Big Data for Precision Medicine"

<http://events.demokritos.gr/?session=big-data-for-precision-medicine>

Wednesday 11 July 2018

6th Hellenic Forum For Science
Technology and Innovation

<http://events.demokritos.gr/>

NCSR "Demokritos", Athens, Greece



Thank you for your attention



<http://project-iasis.eu>



@Project_IASIS