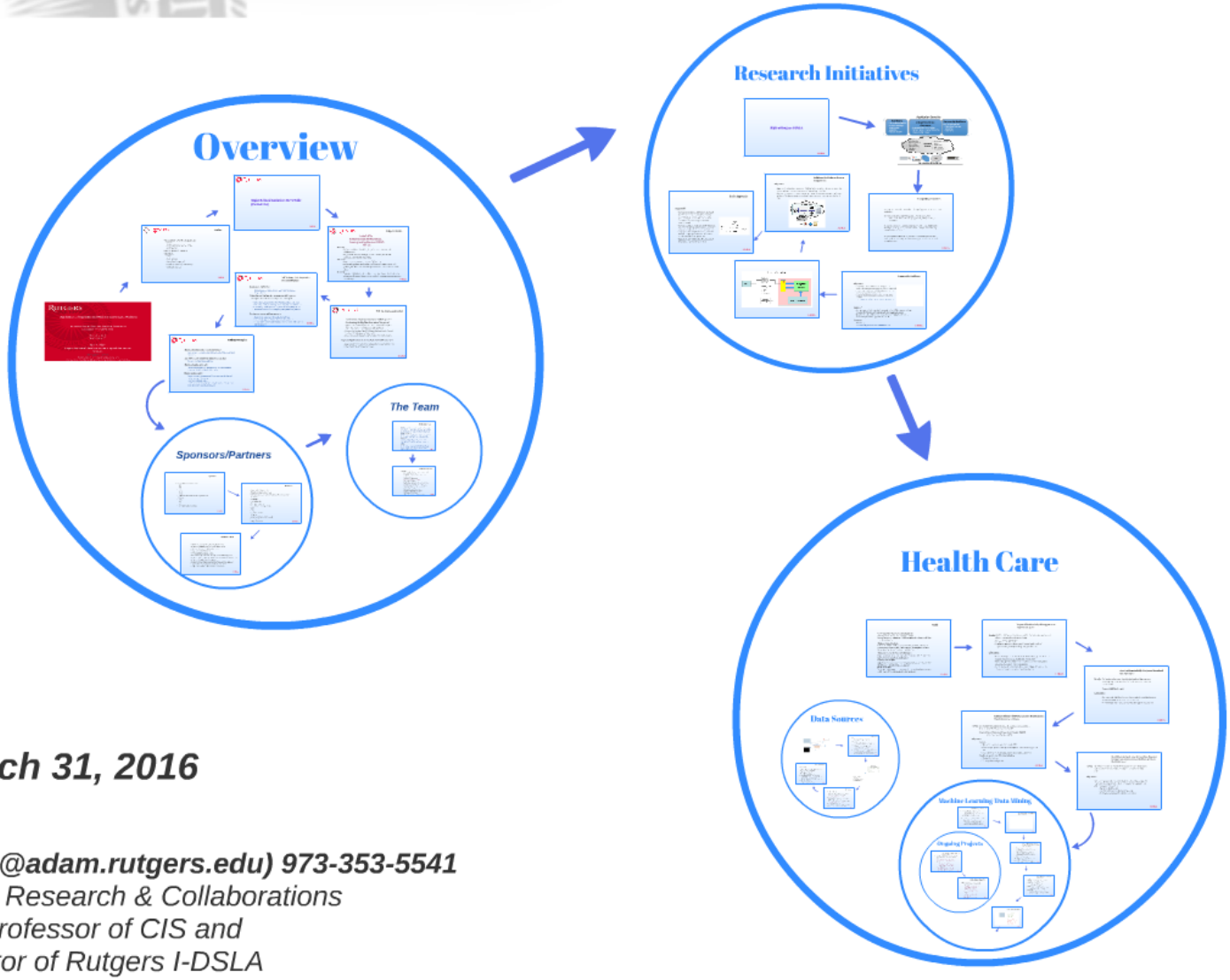
RUTGERS

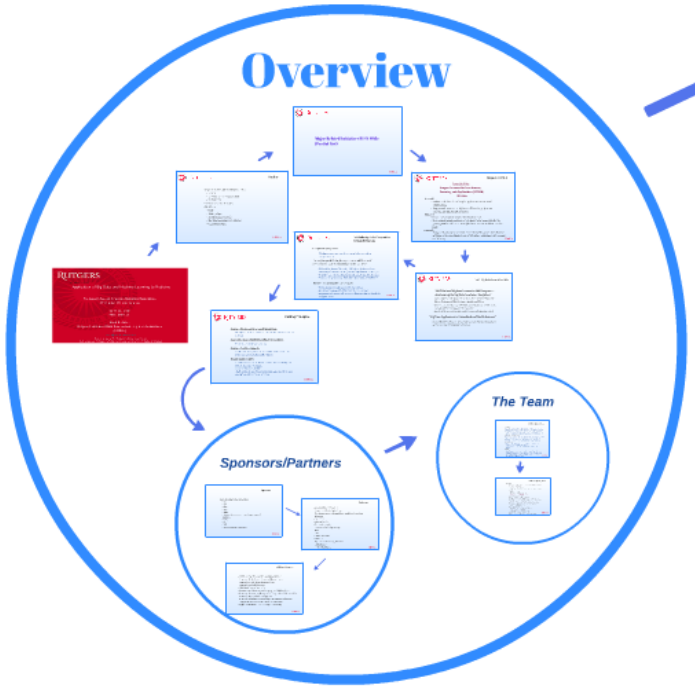
I-DSLA

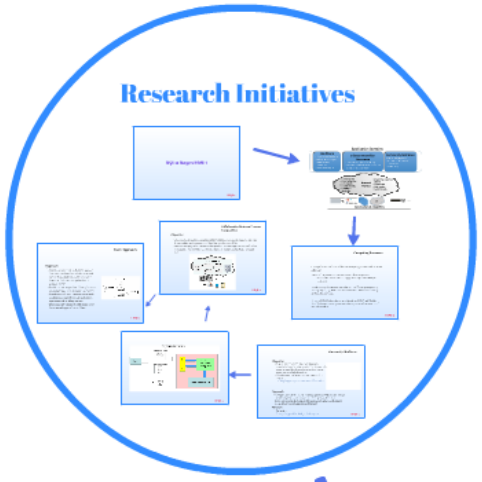


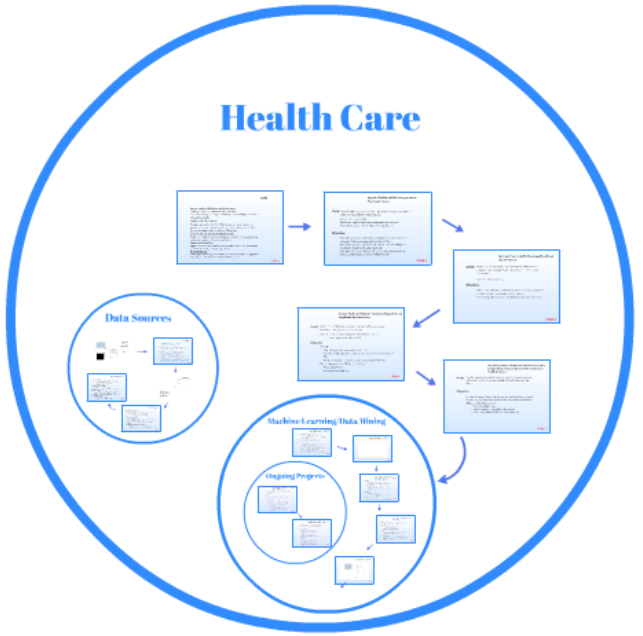
March 31, 2016

Nabil R. Adam (adam@adam.rutgers.edu) 973-353-5541
Vice Chancellor for Research & Collaborations
Distinguished Professor of CIS and
Founding Director of Rutgers I-DSLA

I-DSLA



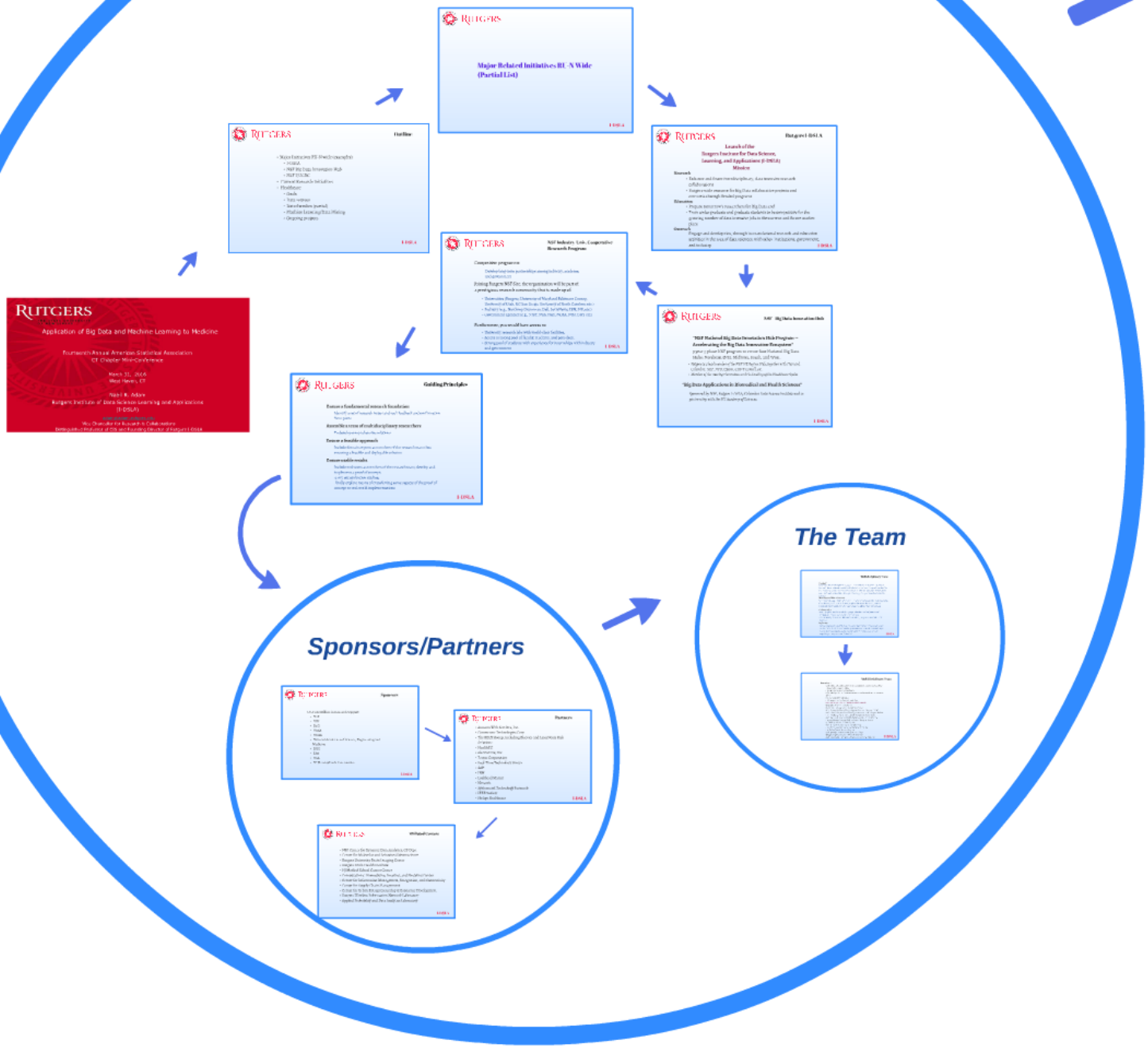




March 31, 2016

Nabil R. Adam (adam@adam.rutgers.edu) 973-353-5541
Vice Chancellor for Research & Collaborations
Distinguished Professor of CIS and
Founding Director of Rutgers I-DSLA

Overview


RUTGERS
Application of Big Data and Machine Learning to Medicine
Major Related Initiatives RU-N Wide (Partial List)
Launch of the Rutgers Institute for Data Science, Learning, and Applications (I-DSLA) Mission
Golding Principles
The Team
Sponsors/Partners

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY

Application of Big Data and Machine Learning to Medicine

Fourteenth Annual American Statistical Association
CT Chapter Mini-Conference

March 31, 2016
West Haven, CT

Nabil R. Adam
Rutgers Institute of Data Science Learning and Applications
(I-DSLA)
adam@adam.rutgers.edu
Vice Chancellor for Research & Collaborations
Distinguished Professor of CIS and Founding Director of Rutgers I-DSLA

RUTGERS

Outline

- Major Initiatives RU-N wide (examples)
  - I-DSLA
  - NSF Big Data Innovation Hub
  - NSF I/UCRC
- Current Research Initiatives
- Healthcare
  - Goals
  - Data sources
  - List of studies (partial)
  - Machine Learning/Data Mining
  - Ongoing projects

RUTGERS

Major Related Initiatives RU-N Wide (Partial List)

I-DSLA

RUTGERS

Rutgers I-DSLA

Launch of the Rutgers Institute for Data Science, Learning, and Applications (I-DSLA)

Mission

Research

- Enhance and foster interdisciplinary, data intensive research collaborations
- Rutgers-wide resource for Big Data collaborative projects and consortia through funded programs

Education

- Prepare tomorrow's researchers for Big Data and
- Train undergraduate and graduate students to be competitive for the growing number of data intensive jobs in the current and future market place

Outreach

Engage and develop ties, through its translational research and education activities in the area of data sciences, with other institutions, government, and industry

I-DSLA

RUTGERS

NSF Industry/Univ. Cooperative Research Program

Competitive program to:

> Develop long-term partnerships among industry, academe, and government

Joining Rutgers NSF Site, the organization will be part of a prestigious research community that is made up of:

- Universities (Rutgers, University of Maryland Baltimore County, University of Utah, UC San Diego, University of North Carolina, etc.)
- Industry (e.g., Northrop Grumman, Dell, LexisNexis, IBM, HP, etc.)
- Government agencies (e.g., NIST, NSA, NAS, NOAA, NIH, DHS etc.)

Furthermore, you would have access to:

- University research labs with world-class facilities,
- Access to strong pool of faculty, students, and post-docs,
- Strong pool of students with experience for internships with industry and government

I-DSLA

RUTGERS

NSF - Big Data Innovation Hub

"NSF National Big Data Innovation Hub Program -- Accelerating the Big Data Innovation Ecosystem"

9 year 3 phase NSF program to create four National Big Data Hubs: Northeast (NE), Midwest, South, and West.

- *Rutgers is a lead member of the NSF NE Region Hub, together with Harvard, Columbia, MIT, NYU, Upenn, CUNY Cornell, etc.*
- *Member of the steering Committee and Co-Leading of the Healthcare Spoke*

"Big Data Applications in Biomedical and Health Sciences"

Sponsored by NSF, Rutgers I-DSLA, Columbia Data Science Institute and in partnership with the NY Academy of Sciences.

I-DSLA

RUTGERS

Guiding Principles

Ensure a fundamental research foundation

Identify a set of research issues and seek feedback and confirmation from peers

Assemble a team of multidisciplinary researchers

To develop comprehensive solutions

Ensure a feasible approach

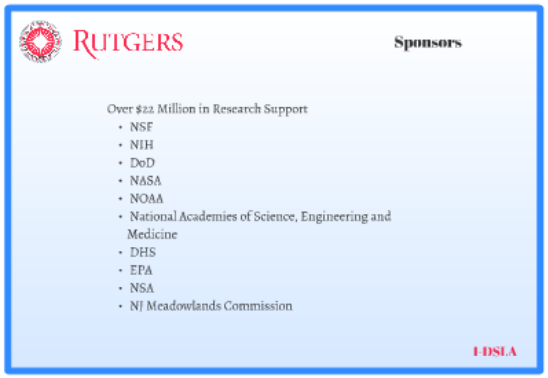
Include domain experts as members of the research team thus ensuring a feasible and deployable solution

Ensure usable results

Include end-users as members of the research team, develop and implement a proof of concept;
carry out evaluation studies;
finally explore means of transferring some aspects of the proof of concept to real-world implementations

I-DSLA

Sponsors/Partners



RUTGERS

Sponsors

Over $22 Million in Research Support

- NSF
- NIH
- DoD
- NASA
- NOAA
- National Academies of Science, Engineering and Medicine
- DHS
- EPA
- NSA
- NJ Meadowlands Commission

I-DSLA

RUTGERS

Partners

- Amazon Web Services, Inc.
- Concurrent Technologies Corp.
- The RELX Group, including Elsevier and LexisNexis Risk Solutions
- HealthEC
- electroCore, Inc.
- Tetrus Corporation
- Real-Time Technology Group
- SAP
- IBM
- Lockheed Martin
- Novartis
- System and Technology Research
- IEEE Society
- Philips Healthcare

I-DSLA

RUTGERS

Affiliated Centers

- NSF Center for Dynamic Data Analytics, CS Dept.
- Center for Molecular and Behavioral Neuroscience
- Rutgers University Brain Imaging Center
- Rutgers Brain Health Institute
- NJ Medical School Cancer Center
- Computational Biomedicine, Imaging, and Modeling Center
- Center for Information Management, Integration, and Connectivity
- Center for Supply Chain Management
- Center for Urban Entrepreneurship & Economic Development
- Rutgers Wireless Information Network Laboratory
- Applied Probability and Data Analytics Laboratory

The Team


RUTGERS
Multidisciplinary Team
Oncology:
Robert Wieder, Prof. of Medicine, Assoc. Director for Clinical and Translational Research, Rutgers New Jersey Medical School Cancer Center, Rutgers Biomedical & Health Sciences; Omar Mahmoud, Asst. Prof. of Radiation Oncology; Ravi Cholesh, Asst. Prof. and Section Chief of Surgical Oncology, Rutgers Biomedical & Health Sciences
Neurology and Neuro Sciences:
Nizar Souayah, Assoc. Prof. and Program Director of Neuromuscular Medicine, MDA clinic director, and Director of EMG/Peripheral Neuropathy Center; Barry R. Komisaruk, Distinguished Prof of Psychology and Adjunct Prof of Radiology.
Cardiovascular:
• Marc Klapholz, Professor and Chair, Dept. of Medicine Chief, Division of Cardiology, Rutgers Biomedical Health Sciences
• John B. Kostis, Director, Cardiovascular Institute, Rutgers Biomedical Health Sciences
Psychiatry:
Petros Levounis, Chair of the Department of Psychiatry at Rutgers New Jersey Medical School and Chief of Psychiatry at University Hospital in Newark, New Jersey; Rashi Aggarwal, Associate Residency Training Director, Dept of Psychiatry at Rutgers NJ Medical School.
I-DSLA


RUTGERS
Multidisciplinary Team
Data Science:
• Nabil Adam, Distinguished Professor of Computer & Info Sys and Founding Director of the Rutgers I-DSLA
• Bob Apel, Professor of Criminal Justice
• Vijay Atluri ,Professor of Computer & Info Sys and Research Director, Rutgers CIMIC
• Valerio Basile, Criminal Justice
• Liz Bonawitz, Asst Professor of Psychology
• Javier Cabrera, Professor of Statistics and Biostatistics
• Chan Choi, Professor of Marketing
• Soon Chun, Professor of Info Sys and Informatics
• Steve Hanson, Professor of Psychology and Director of Rutgers RUBIC
• Michael Katehakis, Professor of Management Science and Info Sys and Chair
• Bart Krekelberg, Professor of Molecular Behavior and Neuroscience
• Michelson, Leslie, Director High Performance & Research Computing
• Dimitris Metaxas, Distinguished Professor to Computer Science
• Joel Miller, Professor of Criminal Justice
• Michale Pavanello, Asst. Professor of Chemistry
• Ronit Stafira, Assistant Professor of Computer & Info Sys
• Pat Shaffo, Chair Professor of Big Data
• Jaideep Vaidya, Professor of Computer & Info Sys
• Gregg Van Ryzin, Professor Public Administration
• Miklos Vasarhelyi, Distinguished Professor Accounting & Info Sys
I-DSLA

RUTGERS

Multidisciplinary Team

Oncology:

Robert Wieder, Prof. of Medicine, Assoc. Director for Clinical and Translational Research, Rutgers New Jersey Medical School Cancer Center, Rutgers Biomedical & Health Sciences; Omar Mahmoud, Asst. Prof. of Radiation Oncology; Ravi Chokshi, Asst. Prof. and Section Chief of Surgical Oncology, Rutgers Biomedical & Health Sciences

Neurology and Neuro Sciences:

Nizar Souayah, Assoc. Prof. and Program Director of Neuromuscular Medicine, MDA clinic director, and Director of EMG/Peripheral Neuropathy Center; Barry R. Komisaruk, Distinguished Prof of Psychology and Adjunct Prof of Radiology.

Cardiovascular:

- Marc Klapholz, Professor and Chair, Dept. of Medicine Chief, Division of Cardiology, Rutgers Biomedical Health Sciences
- John B. Kostis, Director, Cardiovascular Institute, Rutgers Biomedical Health Sciences

Psychiatry:

Petros Levounis, Chair of the Department of Psychiatry at Rutgers New Jersey Medical School and Chief of Psychiatry at University Hospital in Newark, New Jersey; Rashi Aggarwal, Associate Residency Training Director, Dept of Psychiatry at Rutgers NJ Medical School.

I-DSLA

RUTGERS

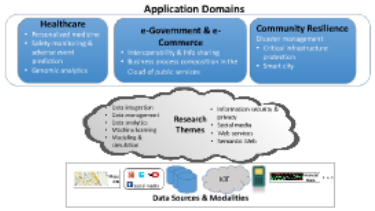
Multidisciplinary Team

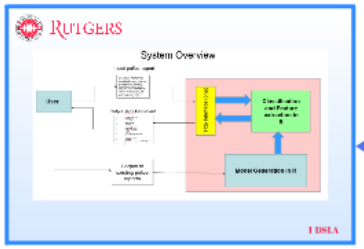
Data Science:

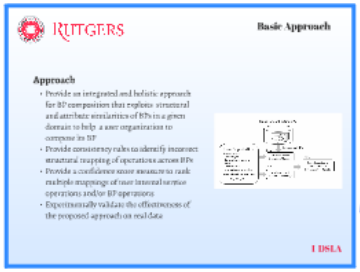
- Nabil Adam, Distinguished Professor of Computer & Info Sys and Founding Director of the Rutgers I-DSLA
- Bob Apel, Professor of Criminal Justice
- Vijay Atluri ,Professor of Computer & Info Sys and Research Director, Rutgers CIMIC
- Valerio Bacak, Criminal Justice
- Liz Bonawitz, Asst Professor of Psychology
- Javier Cabrera, Professor of Statistics and Biostatistics
- Chan Choi, Professor of Marketing
- Soon Chun, Professor of Info Sys and Informatics
- Steve Hanson, Professor of Psychology and Director of Rutgers RUBIC
- Michael Katehakis, Professor of Management Science and Info Sys and Chair
- Bart Krekelberg, Professor of Molecular Behavior and Neuroscience
- Michelson, Leslie, Director High Performance & Research Computing
- Dimitris Metaxas, Distinguished Professor to Computer Science
- Joel Miller, Professor of Criminal Justice
- Michele Pavanello, Asst. Professor of Chemistry
- Basit Shafiq, Assistant Professor of Computer & Info Sys
- Pat Shafto, Chair Professor of Big Data
- Jaideep Vaidya, Professor of Computer & Info Sys
- Gregg Van Ryzin, Professor Public Administration
- Miklos Vasarhelyi, Distinguished Professor Accounting & Info Sys

I-DSLA

Research Initiatives







RUTGERS

R&D at Rutgers I-DSLA

I-DSLA

Application Domains

Healthcare

- Personalized medicine
- Safety monitoring & adverse event prediction
- Genomic analytics

e-Government & e-Commerce

- Interoperability & Info sharing
- Business process composition in the Cloud of public services

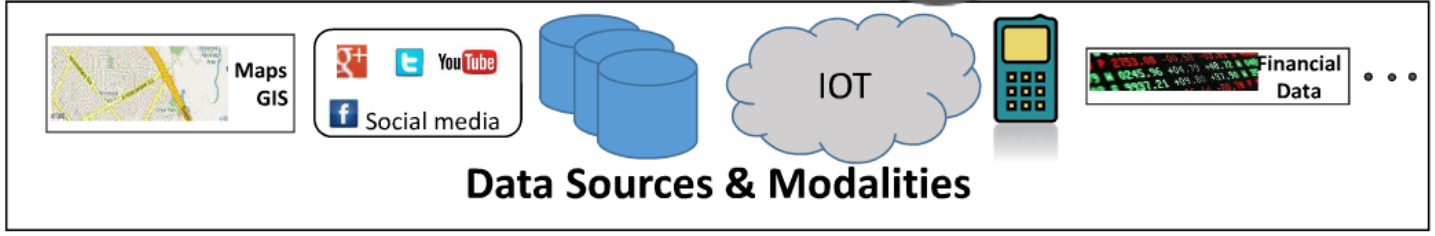
Community Resilience

Disaster management

- Critical infrastructure protection
- Smart city

Research Themes

- Data integration
- Data management
- Data analytics
- Machine learning
- Modeling & simulation
- Information security & privacy
- Social media
- Web services
- Semantic Web



RUTGERS

Computing Resources

A comprehensive and state-of-the-art Computing resource infrastructure, including

A 65 node Linux cluster and two large memory Linux smp servers

- Three of the cluster nodes are equipped with quad NVIDIA Kepler K20 GPUs.

A machine comprises 150 compute nodes and ~750 TB storage supported by both high-speed, e.g., Lustre, and more conventional NFS-mounted storage for home directory space.

A 32 node Dell M610 blade cluster, consisting of two Dell M1000E Modular Blade Enclosures, necessary interconnect/management infrastructure, and supervisory nodes

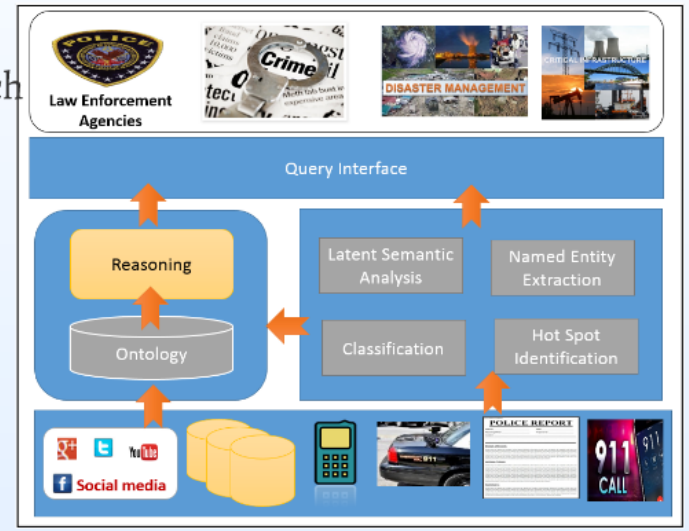
I-DSLA

RUTGERS

Community Resilience

Objectives

- Help analysts at the Law Enforcement Agencies to categorize information using supervised learning approach
- Extract meaningful, relevant information about various entities associated with the incident
- Identify temporal and spatial associations with earlier reports
  - Using integrated queries and semantic Web services



Approach

- Intelligent system built on machine learning algorithms (SVM), semantic ontology, named entity extraction and integration of heterogeneous data sources
- Simple user interface built using R- shiny package and deployed on the internal server with MySQL and MongoDB as the backend

Partners

Tetrus Corp.

Prototype integrated into their production system

I-DSLA

ch

Law Enforcement Agencies

Query Interface

Reasoning

Ontology

Latent Semantic Analysis

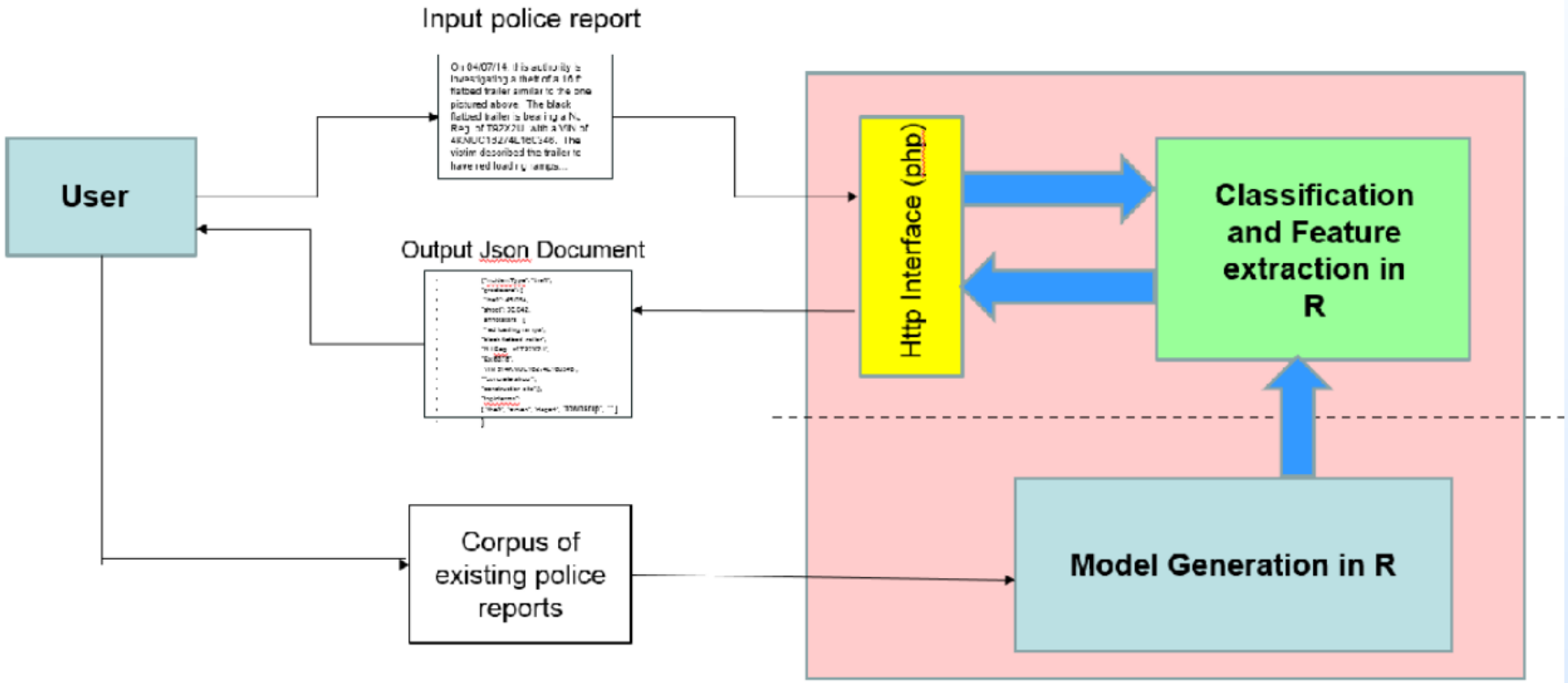
Named Entity Extraction

Classification

Hot Spot Identification

Social media

System Overview

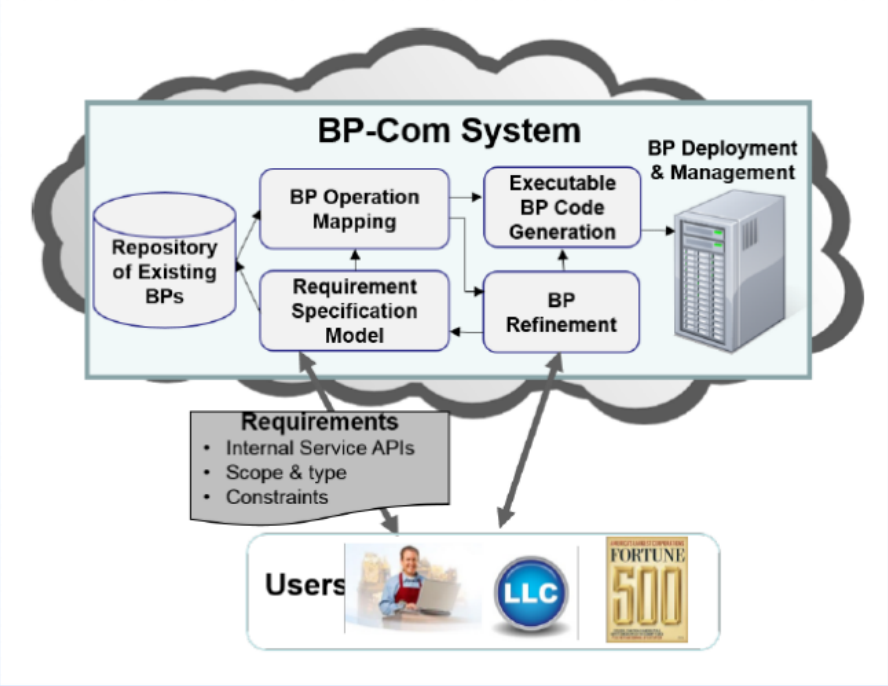


RUTGERS

Collaborative Business Process Composition

Objectives

- Given small and medium enterprises (SMEs) which may not have adequate resources for design, coding, and management of their Business Processes (BPs)
- Compute mapping of the internal service operations of a user organization to collections of related BP -- by exploiting the similarity between structure and attributes of related BPs

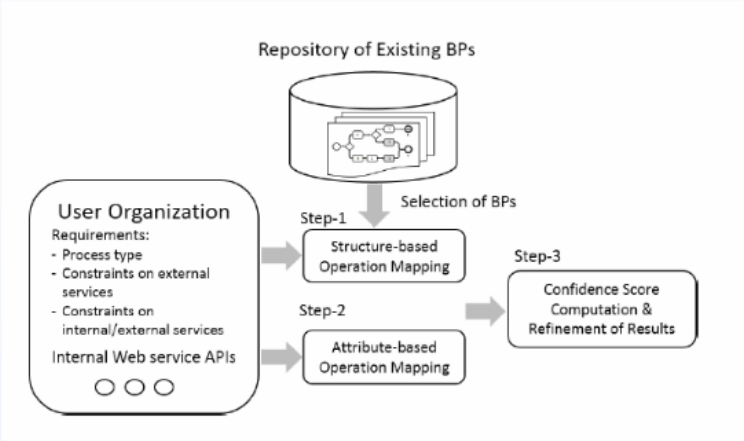


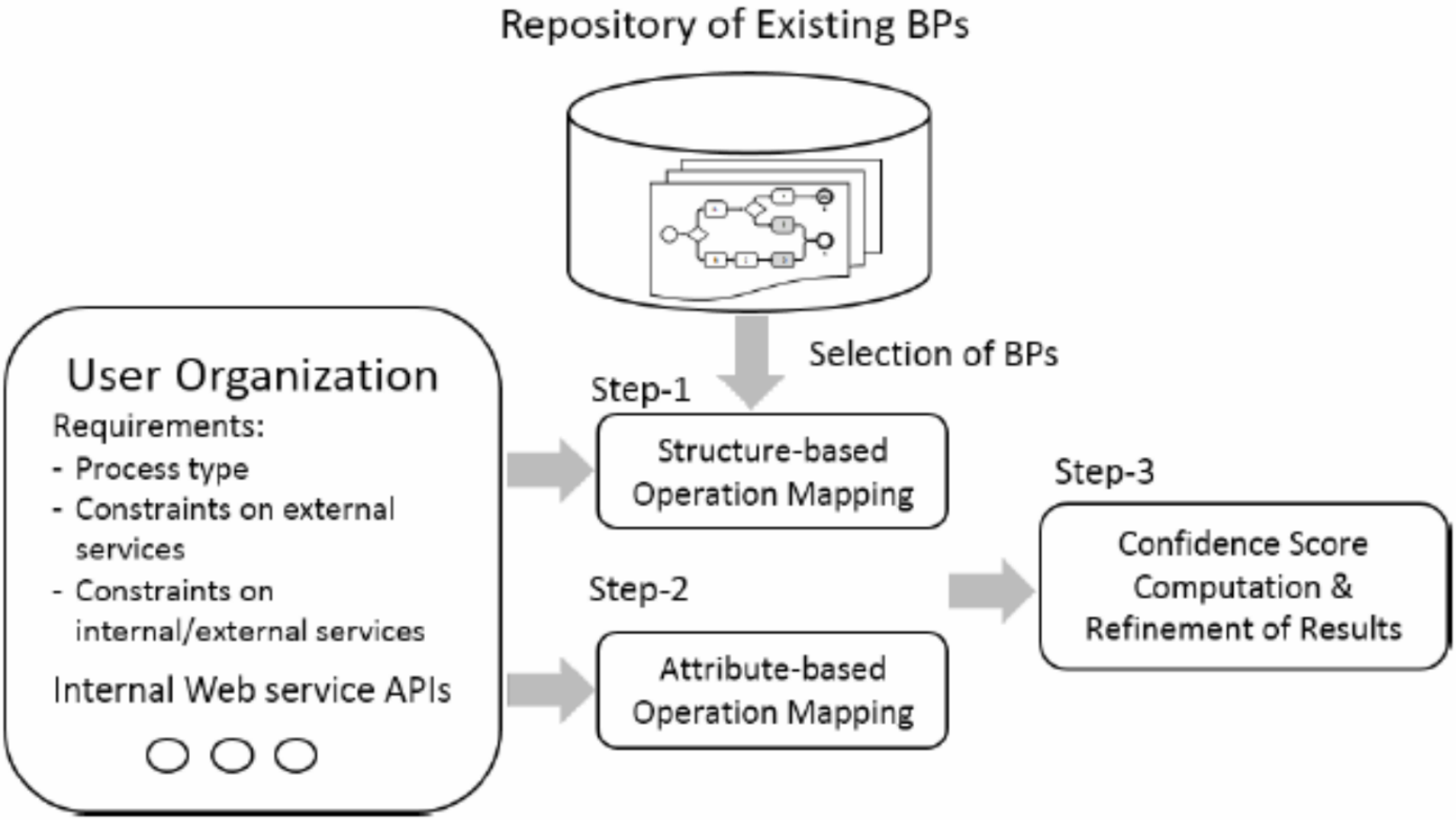
I-DSLA

Basic Approach

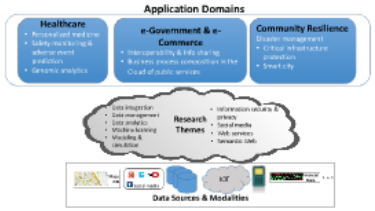
Approach

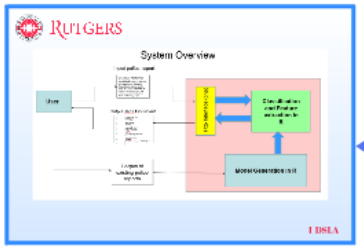
- Provide an integrated and holistic approach for BP composition that exploits structural and attribute similarities of BPs in a given domain to help a user organization to compose its BP
- Provide consistency rules to identify incorrect structural mapping of operations across BPs
- Provide a confidence score measure to rank multiple mappings of user internal service operations and/or BP operations
- Experimentally validate the effectiveness of the proposed approach on real data

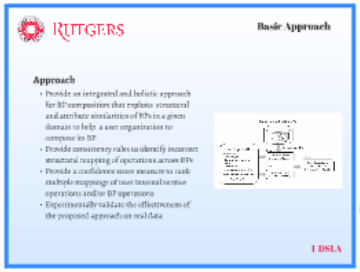



Repository of Existing BPs
Selection of BPs
User Organization
Requirements:
- Process type
- Constraints on external services
- Constraints on internal/external services
Internal Web service APIs
Step-1
Structure-based Operation Mapping
Step-2
Attribute-based Operation Mapping
Step-3
Confidence Score Computation & Refinement of Results

Research Initiatives



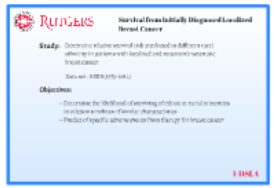


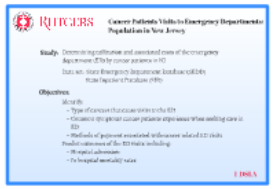


Health Care

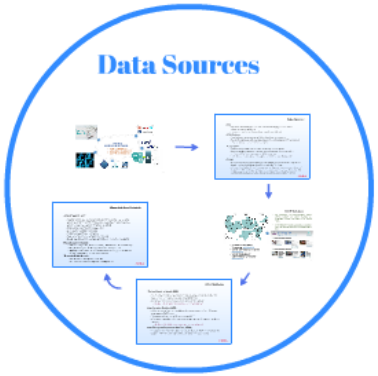

RUTGERS
Goals
I-DSLA

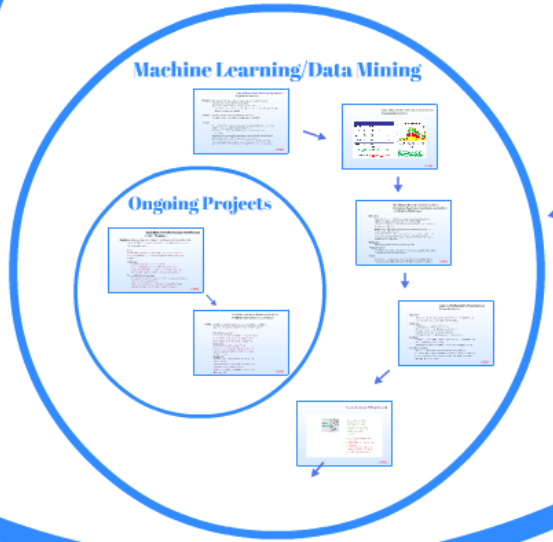

RUTGERS
Impact of Multimorbidity Management on Pancreatic Cancer
I-DSLA


RUTGERS
Survival from Initially Diagnosed Localized Breast Cancer
I-DSLA


RUTGERS
Cancer Patients Visits to Emergency Departments: Population in New Jersey
I-DSLA


RUTGERS
I-DSLA


Data Sources


Machine Learning/Data Mining
Ongoing Projects

RUTGERS

Goals

Improve quality of Healthcare and Reduce Costs

Enabling care to be more preventive and personalized

An estimated savings (according to a McKinsey study) of $300B/yr in the US alone

Focused areas include:

Evidence-based medicine

Combine and analyze a variety of (EHR, financial, operational, clinical, and genomic, etc.) to match treatments with outcomes, predict patients at risk for disease or readmission and provide more efficient care

Device/remote monitoring and Mobile health

Capture and analyze in real-time data from in-hospital and in-home devices, for safety monitoring and adverse event prediction

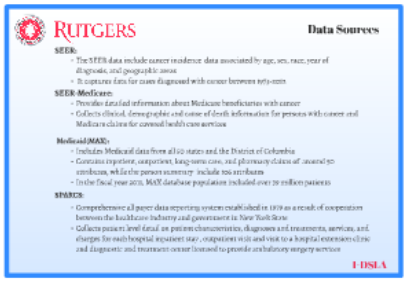
Patient profile analytics

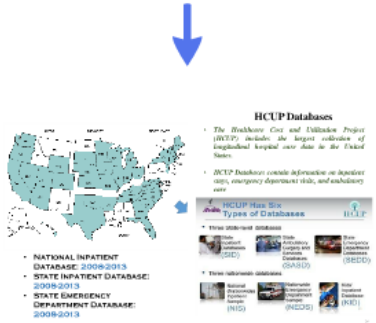
Apply analytics to patient profiles to identify individuals who would benefit from proactive care or lifestyle changes, e.g., diabetes

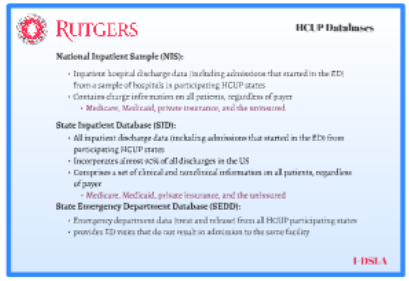
Genomic analytics

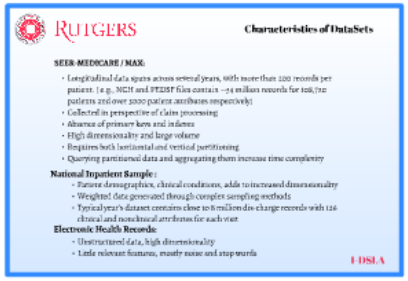
Execute gene sequencing more efficiently and cost effectively and make genomic analysis a part of the regular medical care decision process

Data Sources









DATA INTEGRATION

- *STRUCTURED DATA*
- *SEMI-STRUCTURED DATA*
- *UNSTRUCTURED DATA*

SEER

H·CUP

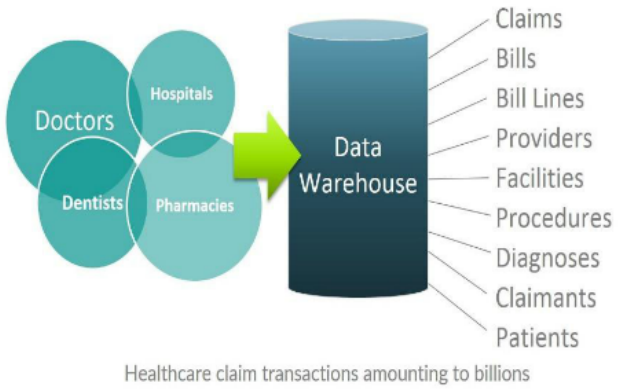
Commission on Cancer
NCDB

YouTube

Social media

Medicaid

Medicare



Healthcare claim transactions amounting to billions

RUTGERS

Data Sources

SEER:

- The SEER data include cancer incidence data associated by age, sex, race, year of diagnosis, and geographic areas
- It captures data for cases diagnosed with cancer between 1973-2012

SEER-Medicare:

- Provides detailed information about Medicare beneficiaries with cancer
- Collects clinical, demographic and cause of death information for persons with cancer and Medicare claims for covered health care services

Medicaid(MAX):

- Includes Medicaid data from all 50 states and the District of Columbia
- Contains inpatient, outpatient, long-term care, and pharmacy claims of around 50 attributes, while the person summary include 106 attributes
- In the fiscal year 2011, MAX database population included over 39 million patients

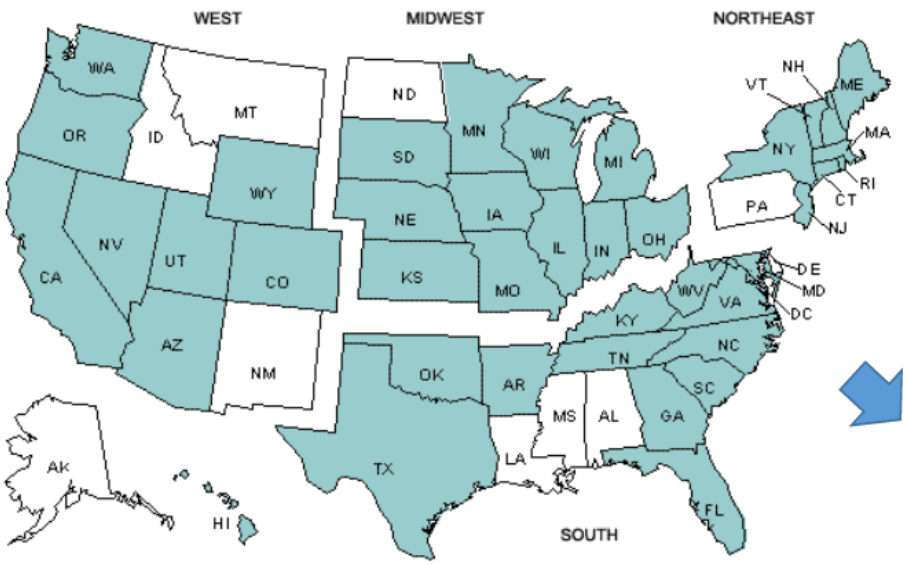
SPARCS:

- Comprehensive all payer data reporting system established in 1979 as a result of cooperation between the healthcare industry and government in New York State
- Collects patient level detail on patient characteristics, diagnoses and treatments, services, and charges for each hospital inpatient stay , outpatient visit and visit to a hospital extension clinic and diagnostic and treatment center licensed to provide ambulatory surgery services

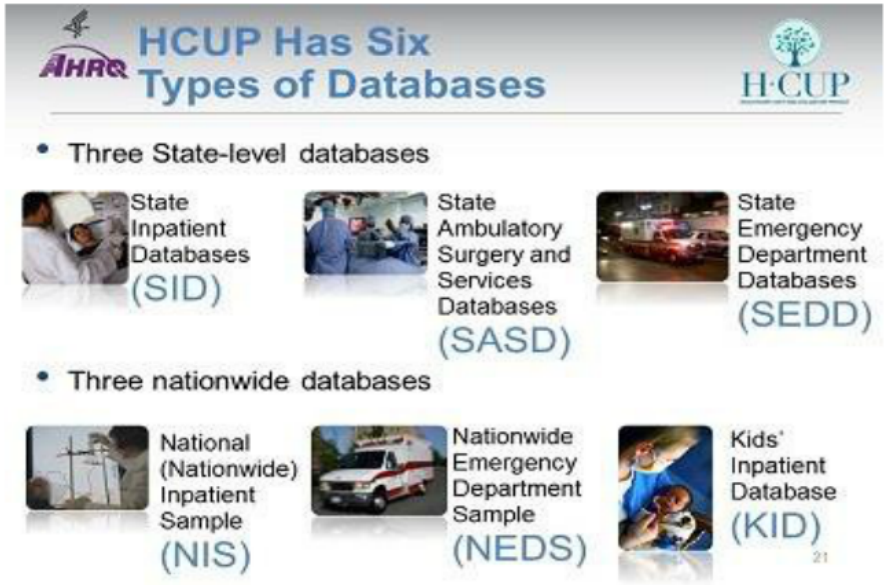
I-DSLA

HCUP Databases

- *The Healthcare Cost and Utilization Project (HCUP) includes the largest collection of longitudinal hospital care data in the United States.*
- *HCUP Databases contain information on inpatient stays, emergency department visits, and ambulatory care*



- **National Inpatient Database: 2008-2013**
- **State Inpatient Database: 2008-2013**
- **State Emergency Department Database: 2008-2013**



RUTGERS

HCUP Databases

National Inpatient Sample (NIS):

- Inpatient hospital discharge data (including admissions that started in the ED) from a sample of hospitals in participating HCUP states
- Contains charge information on all patients, regardless of payer
  - Medicare, Medicaid, private insurance, and the uninsured

State Inpatient Database (SID):

- All inpatient discharge data (including admissions that started in the ED) from participating HCUP states
- Incorporates almost 90% of all discharges in the US
- Comprises a set of clinical and nonclinical information on all patients, regardless of payer
  - Medicare, Medicaid, private insurance, and the uninsured

State Emergency Department Database (SEDD):

- Emergency department data (treat and release) from all HCUP participating states
- provides ED visits that do not result in admission to the same facility

RUTGERS

Characteristics of DataSets

SEER-MEDICARE / MAX:

- Longitudinal data spans across several years, with more than 200 records per patient. (e.g., NCH and PEDSF files contain --34 million records for 108,720 patients and over 3000 patient attributes respectively)
- Collected in perspective of claim processing
- Absence of primary keys and indexes
- High dimensionality and large volume
- Requires both horizontal and vertical partitioning
- Querying partitioned data and aggregating them increase time complexity

National Inpatient Sample :

- Patient demographics, clinical conditions, adds to increased dimensionality
- Weighted data generated through complex sampling methods
- Typical year's dataset contains close to 8 million dis-charge records with 126 clinical and nonclinical attributes for each visit

Electronic Health Records:

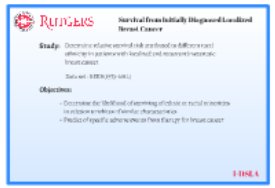
- Unstructured data, high dimensionality
- Little relevant features, mostly noise and stop words

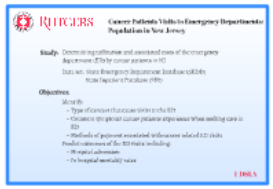
I-DSLA

Health Care

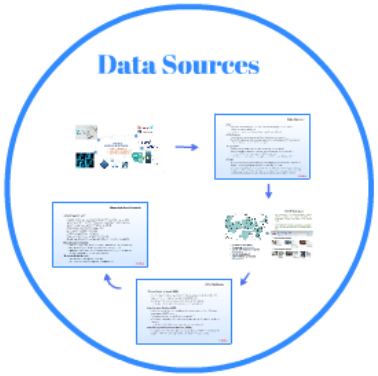

RUTGERS
Goals
I-DSLA

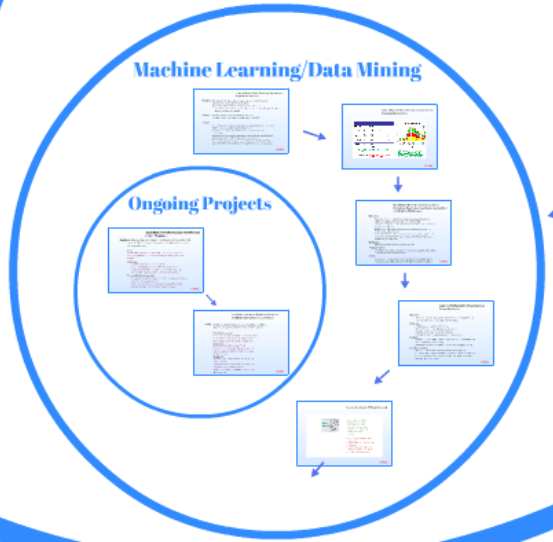

RUTGERS
Impact of Multimorbidity Management on Pancreatic Cancer
I-DSLA


RUTGERS
Survival from Initially Diagnosed Localized Breast Cancer
I-DSLA


RUTGERS
Cancer Patients Visits to Emergency Departments: Population in New Jersey
I-DSLA


RUTGERS
I-DSLA


Data Sources


Machine Learning/Data Mining
Ongoing Projects

RUTGERS

Impact of Multimodality Management on Pancreatic Cancer

Study: Survivability from pancreatic cancer, given various stages, spectrum of adverse events, and combination therapy

- Dataset : SEER –MEDICARE
- Population: patients with associated longitudinal records of hospitalization, prescription, diagnosis, procedure, etc.

Objectives:

- Determine if specific combinations of adjuvant therapy modalities are associated with an increased odds ratio for survival
- Predict the spectrum of toxicities and adverse events resulting from various combinations of cancer treatments
- Determine the characteristics of the patient subpopulations able to tolerate and complete combined modality therapy

I-DSLA

RUTGERS

Survival from Initially Diagnosed Localized Breast Cancer

Study: Determine relative survival risk attributed to different race/ethnicity in patients with localized and recurrent/metastatic breast cancer.

Data set : SEER(1973-2012)

Objectives:

- Determine the likelihood of surviving of ethnic or racial minorities in relation to whites of similar characteristics
- Predict of specific adverse events from therapy for breast cancer

I-DSLA

RUTGERS

Cancer Patients Visits to Emergency Departments: Population in New Jersey

Study: Determining utilization and associated costs of the emergency department (ED) by cancer patients in NJ

Data set: State Emergency Department Database (SEDD)
State Inpatient Database (SID)

Objectives:

Identify:

- Type of cancers that cause visits to the ED
- Common symptoms cancer patients experience when seeking care in ED
- Methods of payment associated with cancer related ED visits

Predict outcomes of the ED visits including:

- Hospital admission
- In hospital mortality rates

I-DSLA

RUTGERS

Identifying Patterns of Comorbid Conditions Prevalent in Major Depression Patients and its Effect on Patient Health Outcomes

Study: Identify patterns of comorbid conditions prevalent in major depression patients and its effect on outcome variables (length of stay and total cost)

Objectives:

Develop various prediction models for patient health outcomes (hospital length of stay and total charges) that include patient demographic and different comorbidity measures:

- Total comorbidity count;
- Specific pattern of comorbid conditions; and
- Count of comorbid condition within each pattern

I-DSLA

Machine Learning/Data Mining


RUTGERS

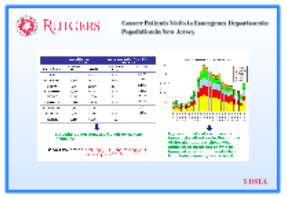
Cancer Patients' Visits to Emergency Departments: Population in New Jersey

Objectives: (a) Characterize the disposition of the acute patient setting, ED boarders, present coding realities, payer, clinical conditions.
(b) Identify significant differences among patient populations:
- more likely admitted to hospital from ED (as opposed to treated and released from ED)
- had longer admitted to hospital

Analyses: Initial case problem - preprocessing and modeling
Predictive models - unconditional logistic regression and ANN

Results:
- Most common cancer diagnoses were: lung (prim), leukemia (s.t. chronic myeloid), lymphoma (s.t. Hodgkin), thyroid primaries (+ 0%)
- Oral Cancer patients were male and Hispanic, Non-Hispanic, or Unknown
- Both African American and Hispanic patients had significantly lower likelihoods of being treated in the EMT-ACP compared to the Caucasians
- Patients in the extreme age groups were more likely to statistically significant to receive EMT-ACP or Protocol Evaluation in comparison to the other age groups

I-DSLA


RUTGERS

Cancer Patients' Visits to Emergency Departments: Population in New Jersey

I-DSLA


RUTGERS

Identifying Patterns of Comorbid Conditions Prevalent in Major Depression Patients and Its Effect on Patient Health Outcomes

Objectives:
- Clustering comorbid conditions into homogeneous subgroups taking into account the interaction between them, primary diagnosis
- Using a bi-clustering approach, further identification of unique patient cohorts from homogeneous subgroups
- Study the effect of the above techniques on various measures: re-admission, length of stay, and total charges
- Evaluate the efficiency of the proposed methodology (in terms of identification of methods that were unsuccessful) relative to expert knowledge measures, e.g., Charlson's Comorbidity Index

Challenges:
- Clustering high dimensional, weighted categorical data

Proposed Solutions:
- Development weighted cluster ing algorithms to allow some categorical variables and observations to contribute more strongly to cluster solutions than others

Results:
- Implementation of unique outcomes can be achieved by satisfying the cluster indicators and considering cause-driving risk factors, demographic variables

I-DSLA

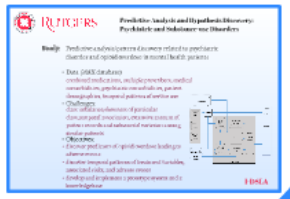

RUTGERS

Application of Precision Medicine to Understand Cancer Phenotypes

Study: Application of big data concepts to precision medicine to understand causes of higher frequencies of occurrence of aggressive cancers in minority populations

- Data
  - Genomic data, electronic health records, vaccine banks, claims data, drug use, procedures, prescription drugs, hospital encounters, ICD9 diagnoses
- Challenges
  - Algorithms need to be non-distance based
  - Non-Euclidean geometry, as Complexity of genomic datasets
  - Large number of variables, much larger than the sample size
  - Analytical value of datasets of heterogeneous large scale data
- Proposed Solutions (ongoing)
  - Use of clustering for revealing patient fully relevant combinations of clinical and genomic risk factors
  - Novel computation of a general tool — discussion drop reduction, biological reinforcement, validation, and eventual discovery

I-DSLA


RUTGERS

Predictive Analysis and Hypothesis Discovery: Psychiatric and Substance-use Disorders

Study: Predictive analysis/pattern discovery related to psychiatric disorders and opioid overdose in mental health patients

- Data (SASK database)
  - combined medications, multiple procedures, medical comorbidities, psychiatric comorbidities, patient demographics, hospital patterns of service use
- Challenges
  - Data collection and retrieval of particular characteristics, patient prescription, procedure codes, of patients who developed substance use outcomes among similar patients
- Objectives:
  - Discover predictors of opioid overdose leading to adverse events
  - Describe temporal patterns of diseases and variables, associated with, and adverse events
  - Develop and implement a preemptive system and a knowledgebase

I-DSLA


RUTGERS

Impact of Multimorbidity Management on Pancreatic Cancer

Objectives:
- Determine if specific combinations of chronic therapy medication are associated with an increased odds ratio for survival

Challenges:
- Highly dispersed and unstructured data
- Control bias (for perspective of data representing)
- Large globules of data spaces across retrieved years
- Unstructured and mixed partitioning of the large datasets

Analyses:
- Logistic Regression (robust, ordinal outcome variables, and pseudo-unity)
- Propensity scoring and matching
- Cox Proportional Hazards Model (for survival in relative long-term use)

Results (ongoing):
- So far the results show: patients taken more only adjuvant chemotherapy to higher likelihood of survival than those with unresected tumors and chemotherapy and radiation and adjuvant radiation therapy to (over 18 months)

I-DSLA


RUTGERS

Next Gen Healthcare: EHRs for Research

I-DSLA

Ongoing Projects

RUTGERS

Application of Precision Medicine to Understand Cancer Phenotypes

Study: Application of big data concepts to precision medicine to understand causes of higher frequencies of occurrence of aggressive cancers in minority populations

- Data
  - Genomic data, electronic health records, vaccine banks, claims data, drug use, procedures, prescription drugs, hospital encounters, ICD9 diagnoses
- Challenges
  - Algorithms need to be non-distance based
  - Non-Euclidean geometry, as Complexity of genomic datasets
  - Large number of variables, much larger than the sample size
  - Analytical value of datasets of heterogeneous large scale data
- Proposed Solutions (ongoing)
  - Use of clustering for revealing patient fully relevant combinations of clinical and genomic risk factors
  - Novel computation of a general tool — discussion drop reduction, biological reinforcement, validation, and eventual discovery

I-DSLA

RUTGERS

Predictive Analysis and Hypothesis Discovery: Psychiatric and Substance-use Disorders

Study: Predictive analysis/pattern discovery related to psychiatric disorders and opioid overdose in mental health patients

- Data (SASK database)
  - combined medications, multiple procedures, medical comorbidities, psychiatric comorbidities, patient demographics, hospital patterns of service use
- Challenges
  - Data collection and retrieval of particular characteristics, patient prescription, procedure codes, of patients who developed substance use outcomes among similar patients
- Objectives:
  - Discover predictors of opioid overdose leading to adverse events
  - Describe temporal patterns of diseases and variables, associated with, and adverse events
  - Develop and implement a preemptive system and a knowledgebase

I-DSLA

RUTGERS

Cancer Patients Visits to Emergency Departments: Population in New Jersey

Objectives: (a) Characterize the disposition of the cancer patient visiting ED based on: patient demographics, payer, clinical conditions
(b) Identify significant differences among patient populations :
ones directly admitted to hospital from ED and ones who were discharged from ED but later get admitted to hospital

Analysis: Imbalance problem : propensity scoring and matching
Prediction models : unconditional logistic regression and SVM

Results:

- Most common cancer diagnoses were: lung (30%), leukemia (12.4%), colon (11.72%), lymphoma(11.6%), breast (7.3%) and prostate (7.3%)
- Chief complaints: pain, malaise and fatigue, fever, syncope, respiratory abnormalities
- Both African American and Hispanic patients had significantly lower likelihoods of being transferred to SNF /ICF compared to the Caucasians
- Patients in the 65-75 year age group were more likely (statistically significant) to utilize SNF/ICF or Home Healthcare Services compared to the other age groups

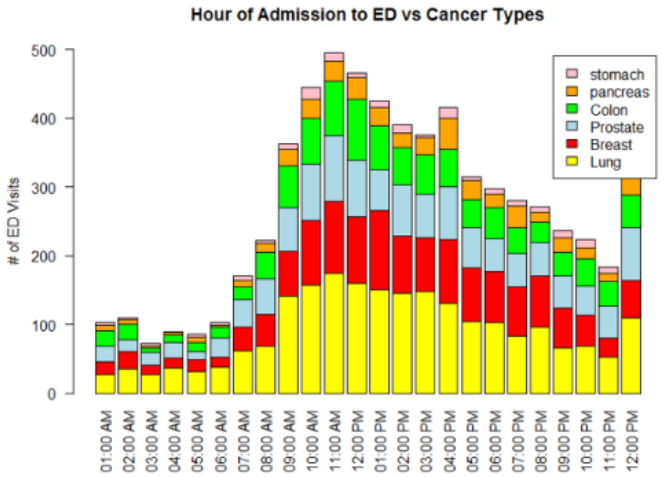
I-DSLA

Cancer Patients Visits to Emergency Departments: Population in New Jersey

	Cases of ED Visits (N = 37,080)		Cancer Cases in New Jersey (2015) (N = 51,410)	
Type of Cancer	Number of Visits	% of Total	Number of New Cases	% of Total
LUNG	11,115	29.98%	5,830	11.34%
LEUKEMIA	4,589	12.38%	1,610	3.13%
COLON	4,345	11.72%	4,260	8.29%
LYMHOMA	4,294	11.58%	2,310	4.49%
BREAST	2,706	7.30%	7,310	14.22%
PROSTATE	2,688	7.25%	7,270	14.14%
PANCREATIC	2,245	6.05%	NA	NA
UTERINE	1,878	5.06%	2,260	4.40%
OTHERS	3,220	8.68%	NA	NA

ED visits do not necessarily mirror cancer incidence

Model Validation: training (75%) and testing (25%)
accuracy = 90.23%



legitimate patients who needs immediate attention in the middle of the night versus those who, either seek narcotics or whose financial situation prevents them from being seen by a specialist

I-DSLA

RUTGERS

Identifying Patterns of Comorbid Conditions Prevalent in Major Depression Patients and its Effect on Patient Health Outcomes

Objectives:

- Clustering comorbid conditions into homogeneous subgroups taking into account the occurrence by chance (using Somer's D statistic)
- Using a bi clustering approach, further identification of unique patterns in these homogeneous subgroups
- Study the effect of the cluster indicators on various outcome variables (e.g., Length of Stay and Total Charge)
- Evaluate the efficiency of the proposed methodology in contrast to traditional methods that uses numerical indices to represent comorbidity measure, e.g., Charlson's Comorbidity Index

Challenges:

Clustering high dimensional, weighted sampled data

Proposed Solution:

Developed weighted clustering algorithms that takes into account sample weight and also weight assigned to the variables

Results:

Better prediction of patient outcome can be achieved by including the cluster indicators and comorbidity count along with patient demographic variables

I-DSLA

RUTGERS

Impact of Multimodality Management on Pancreatic Cancer

Objective:

Determine if specific combinations of adjuvant therapy modalities are associated with an increased odds ratio for survival

Challenges:

- Highly disparate and unstructured data
- Created from the perspective of claim processing
- Longitudinal data spans across several years
- Horizontal and vertical partitioning of the large datasets

Analysis:

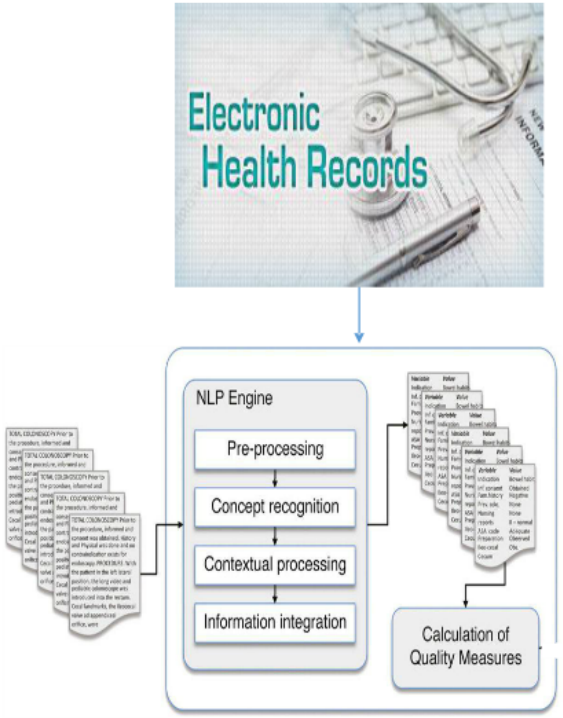
- Logistic Regression (robust, ordinal outcome variables, and conditional)
- Propensity scoring and matching
- Cox Proportional Hazards Model for survival Analysis (log-rank test)

Results (ongoing):

So far the results show patients administered only adjuvant chemotherapy has higher likelihood of survival than those administered combination of chemotherapy and radiation and adjuvant radiation therapy in the early months

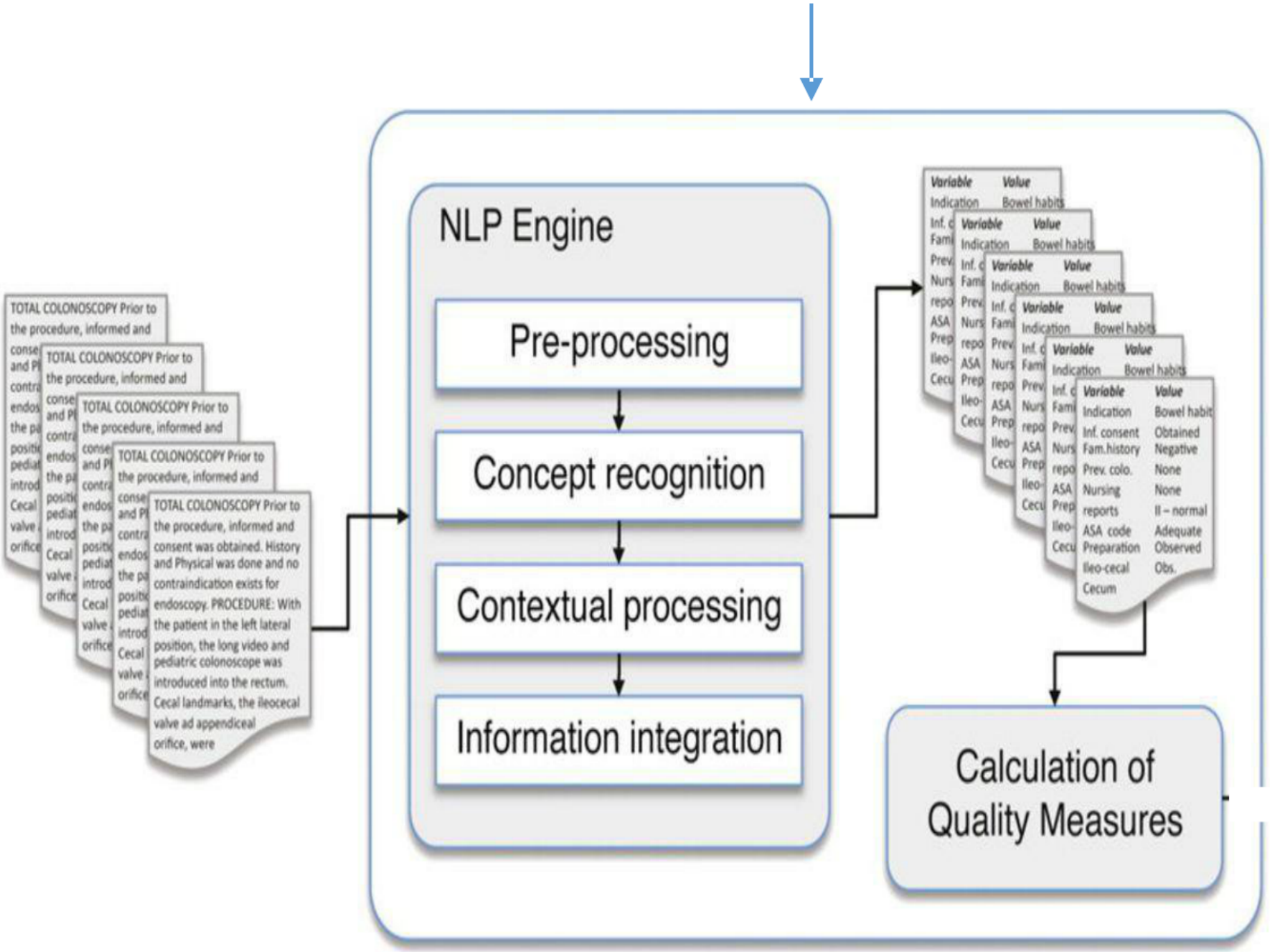
I-DSLA

NextGen Healthcare: EHRs for Research



Similar system has been deployed in the field of community resilience, helping crime analysts to process crime reports and support advanced analytics

- Classification : Support vector Machines
- Topic Modeling : Latent Dirichlet Allocation
- Feature Selection and Ranking: Recursive Feature Elimination
- Named Entity Recognition


TOTAL COLONOSCOPY Prior to the procedure, informed and consent was obtained. History and Physical was done and no contraindication exists for endoscopy. PROCEDURE: With the patient in the left lateral position, the long video and pediatric colonoscope was introduced into the rectum. Cecal landmarks, the ileocecal valve ad appendiceal orifice, were
NLP Engine
Pre-processing
Concept recognition
Contextual processing
Information integration
Variable Value
Indication Bowel habits
Inf. consent Obtained
Fam.history Negative
Prev. colo. None
Nursing reports None
ASA code II – normal
Preparation Adequate
Ileo-cecal Observed
Cecum Obs.
Calculation of Quality Measures

Ongoing Projects

RUTGERS

Application of Precision Medicine to Understand Cancer Phenotypes

Study: Application of big data concepts to precision medicine to understand causes of higher frequencies of occurrence of aggressive cancers in minority populations

- Data
  Genomic data, electronic health records, social media, claims data (diagnosis, procedures, prescription drugs, hospitalizations), HCUP databases
- Challenges
  - Algorithms tend to be non-deterministic
  - Size, heterogeneity, and complexity of genomic datasets
  - Large number of variables, much larger than the sample size
  - Analytics/visualization of heterogeneous large scale data
- Proposed Solutions(ongoing)
  - Use of clustering for revealing potentially relevant combinations of clinical and genomic risk factors
  - Novel computational approaches – dimensionality reduction, biological enrichment, validation, rare event discovery

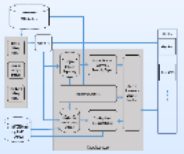
I-DSLA

RUTGERS

Predictive Analysis and Hypothesis Discovery: Psychiatric and Substance-use Disorders

Study: Predictive analysis/pattern discovery related to psychiatric disorder and opioid overdose in mental health patients

- Data (MAX database)
  combined medications, multiple prescribers, medical comorbidities, psychiatric comorbidities, patient demographics, temporal patterns of service use
- Challenges:
  class imbalance,skewness of particular class,temporal association, extensive amount of patient records and substantial variation among similar patients
- Objectives:
- discover predictors of opioid overdose leading to adverse events
- discover temporal patterns of treatment variables, associated risks, and adverse events
- develop and implement a prototype system and a knowledgebase



I-DSLA

RUTGERS

NextGen Healthcare: EHRs

Similar system has been

RUTGERS

Application of Precision Medicine to Understand Cancer Phenotypes

Study: Application of big data concepts to precision medicine to understand causes of higher frequencies of occurrence of aggressive cancers in minority populations

- Data
  Genomic data, electronic health records, social media, claims data (diagnosis, procedures, prescription drugs, hospitalizations), HCUP databases
- Challenges
  - Algorithms tend to be non-deterministic
  - Size, heterogeneity, and complexity of genomic datasets
  - Large number of variables, much larger than the sample size
  - Analytics/visualization of heterogeneous large scale data
- Proposed Solutions(ongoing)
  - Use of clustering for revealing potentially relevant combinations of clinical and genomic risk factors
  - Novel computational approaches -- dimensionality reduction, biological enrichment, validation, rare event discovery

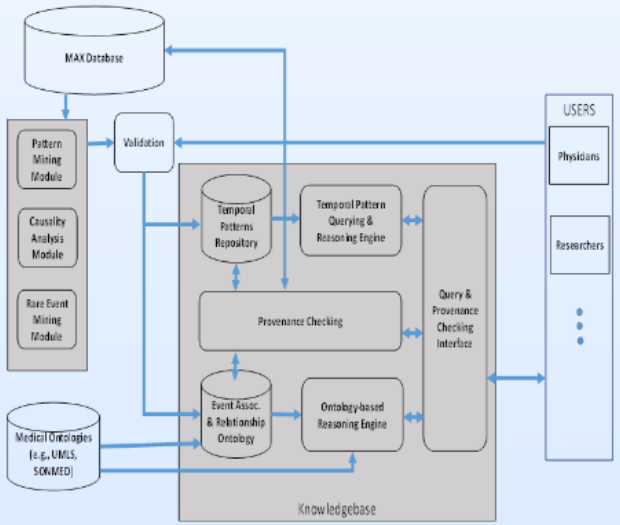
I-DSLA

RUTGERS

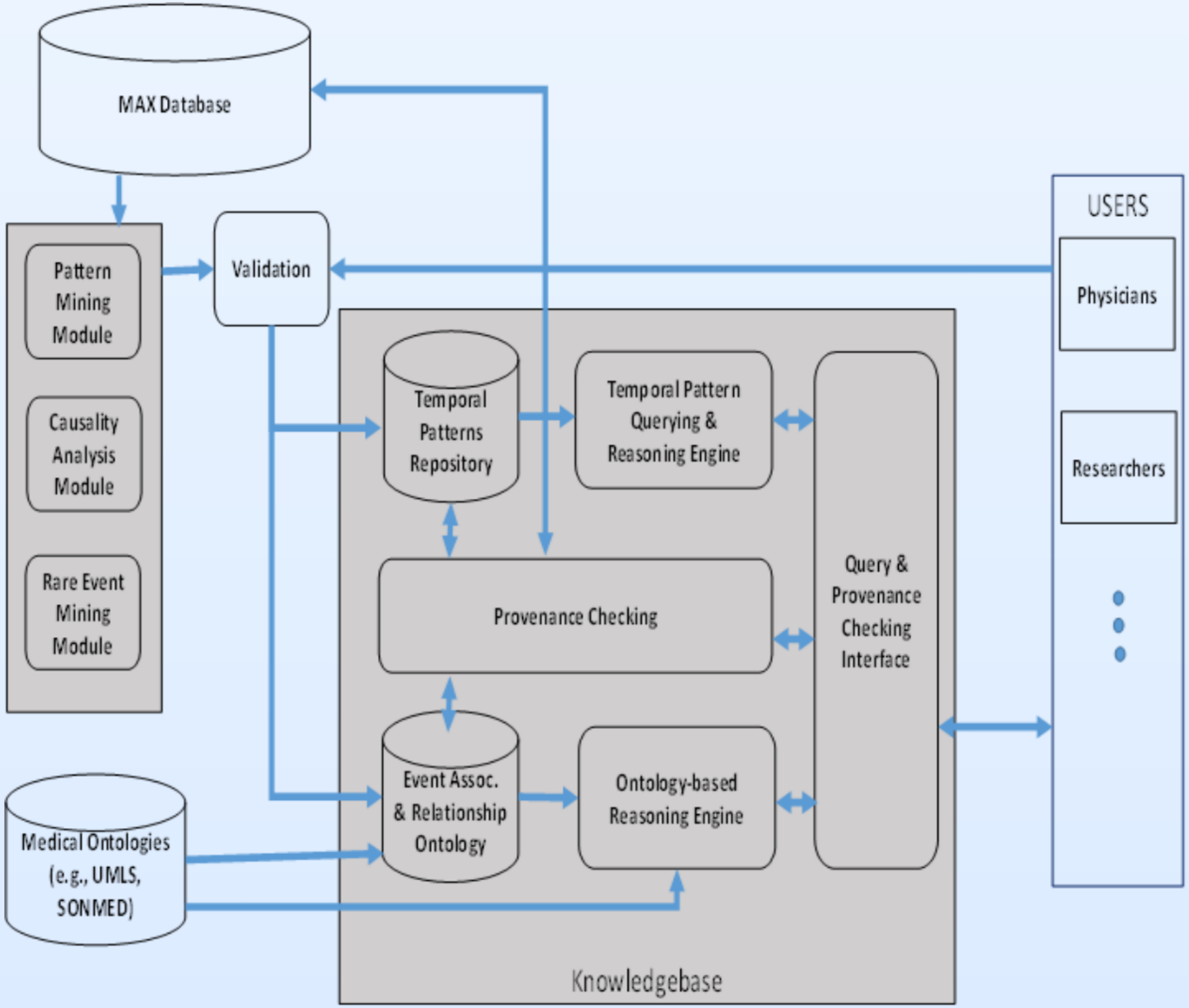
Predictive Analysis and Hypothesis Discovery: Psychiatric and Substance-use Disorders

Study: Predictive analysis/pattern discovery related to psychiatric disorder and opioid overdose in mental health patients

- Data (MAX database)
  combined medications, multiple prescribers, medical comorbidities, psychiatric comorbidities, patient demographics, temporal patterns of service use
- Challenges:
  class imbalance,skewness of particular class,temporal association, extensive amount of patient records and substantial variation among similar patients
- Objectives:
- discover predictors of opioid overdose leading to adverse events
- discover temporal patterns of treatment variables, associated risks, and adverse events
- develop and implement a prototype system and a knowledgebase

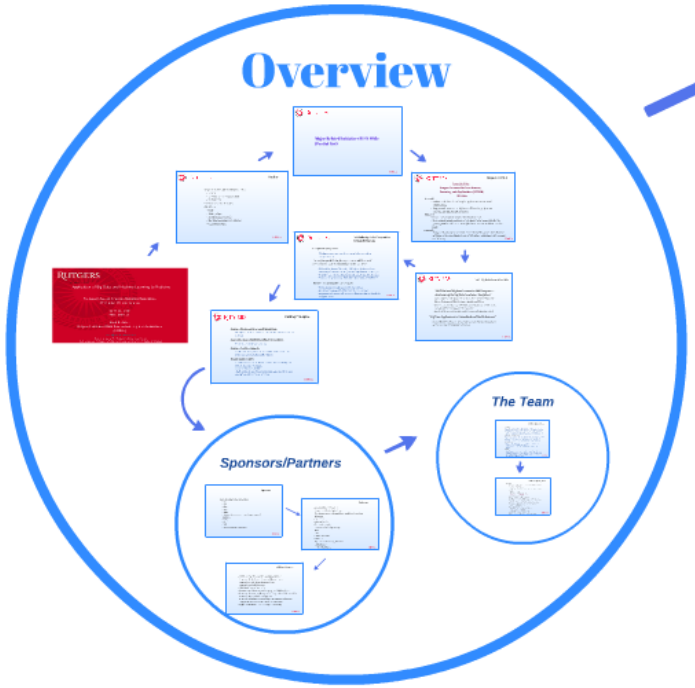


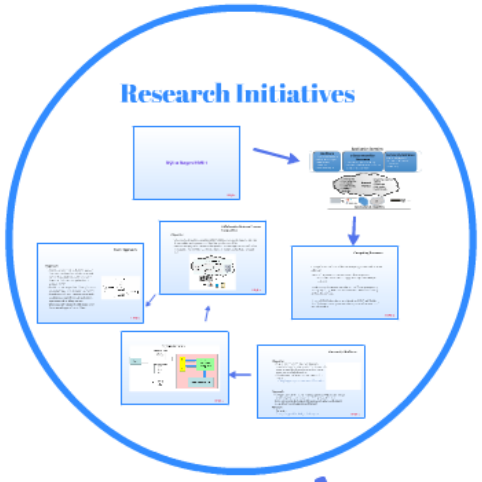
I-DSLA

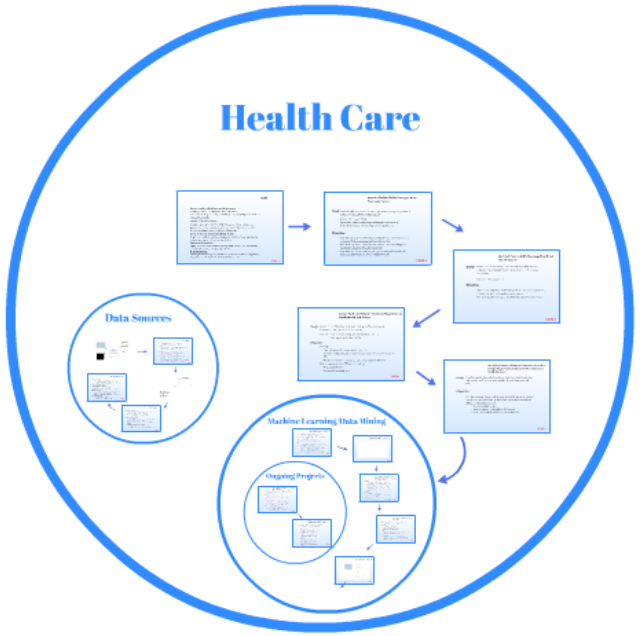

MAX Database
Pattern Mining Module
Causality Analysis Module
Rare Event Mining Module
Validation
Temporal Patterns Repository
Temporal Pattern Querying & Reasoning Engine
Provenance Checking
Event Assoc. & Relationship Ontology
Ontology-based Reasoning Engine
Query & Provenance Checking Interface
Medical Ontologies (e.g., UMLS, SONMED)
Knowledgebase
USERS
Physicians
Researchers

es,

I-DSLA







March 31, 2016

Nabil R. Adam (adam@adam.rutgers.edu) 973-353-5541
Vice Chancellor for Research & Collaborations
Distinguished Professor of CIS and
Founding Director of Rutgers I-DSLA