Data & Analytics Council

EHRs as a Data Source

Friday, June 20, 2014

11:00am-12:00 pm ET

Reminder:

- Please mute your line when not speaking (* 6 to mute, *7 to unmute)

- This call is being recorded

Agenda

- Welcome and introduction

- Speakers
  - Charles Boicey, Enterprise Analytics Architect, Stony Brook Medicine
  - Nitesh Chawla, PhD, Director of The Interdisciplinary Center for Network Science & Applications, University of Notre Dame
  - Simon Beaulah, Director Healthcare Strategy, Linguamatics

- General Discussion

eHEALTH INITIATIVE
Real Solutions. Better Health.

Saritor: A Healthcare Data Ecosystem to Advance Clinical Practice and Research

Charles Boicey, MS, RN-BC, CPHIMS

Enterprise Analytics Architect

Stony Brook Medicine

Forget What You Know: Jacob Barnett

http://youtu.be/Uq-FOOQ1TpE

Why Saritor?

- New sources of health data are emerging that are not handled well by traditional BI/data storage

- The volume, complexity, diversity, & timeliness of healthcare data is rapidly increasing

- Patients are gaining much more insight and interest in managing their own health

- Need for Predictive/Prescriptive Analytics to support pro-active healthcare paradigm

Limits of the Current Model

- The Electronic Medical Record is not designed to process high volume/velocity data, nor is it intended to handle complex operations such as anomaly detection, machine learning, building complex algorithms or pattern set recognition.

- Enterprise Data Warehouses suffer from a latency factor of up to 24 hours. The EDW serves clinicians, operations, quality and research retrospectively as opposed to in real time.

2010 - 212
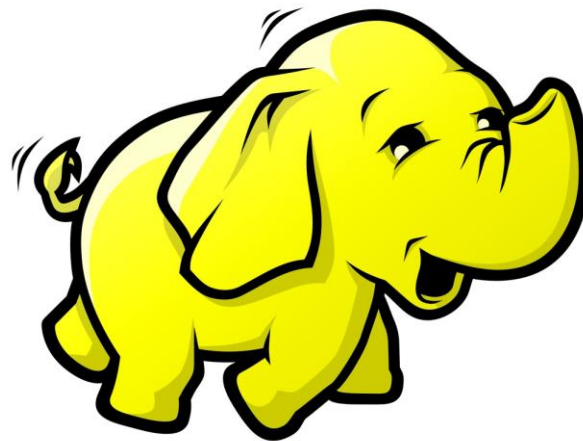
Big Data = Complete Data

- The Electronic Medical Record is primarily transactional taking feeds from source systems via an interface engine.

- The Enterprise Data Warehouse is a collection of data from the EMR and various source systems in the enterprise.

- In both cases decisions are made concerning data acquisition.

- A Big Data system is capable of ingesting and storing healthcare data in total and in real time.

Saritor Data Sources

- Legacy Systems
  - Print to Text or Delimited String
- All HL7 Feeds (EMR source systems)
- All EMR Initiated Data (Stored Procedures)
- Device Data (in one minute intervals)
  - Physiological Monitors (HL7)
  - Ventilators (HL7)
  - Smart Pumps
- Social Media (POC)
  - Healthcare Organization Sentiment Analysis
  - Patient Engagement
- Home Monitoring (POC)
- Real Time Location System (RFID)
- Hospital Sensors
- Genomic Data

Saritor Initial Functionality

- Ingestion of legacy EMR data (20 years)
- Integration with EMR to view legacy data
- Integration with UCI analytics platform (Tableau)
- 30 Day Readmit Prediction (UCI Centric)
- Early Sepsis Detection & Notification
- Rapid Response Team Deployment
- Home Monitoring Analytics
  - Fitbit
  - SyncMetrics
- Social Media Sentiment Analysis

Saritor: A Modern Healthcare Data Platform
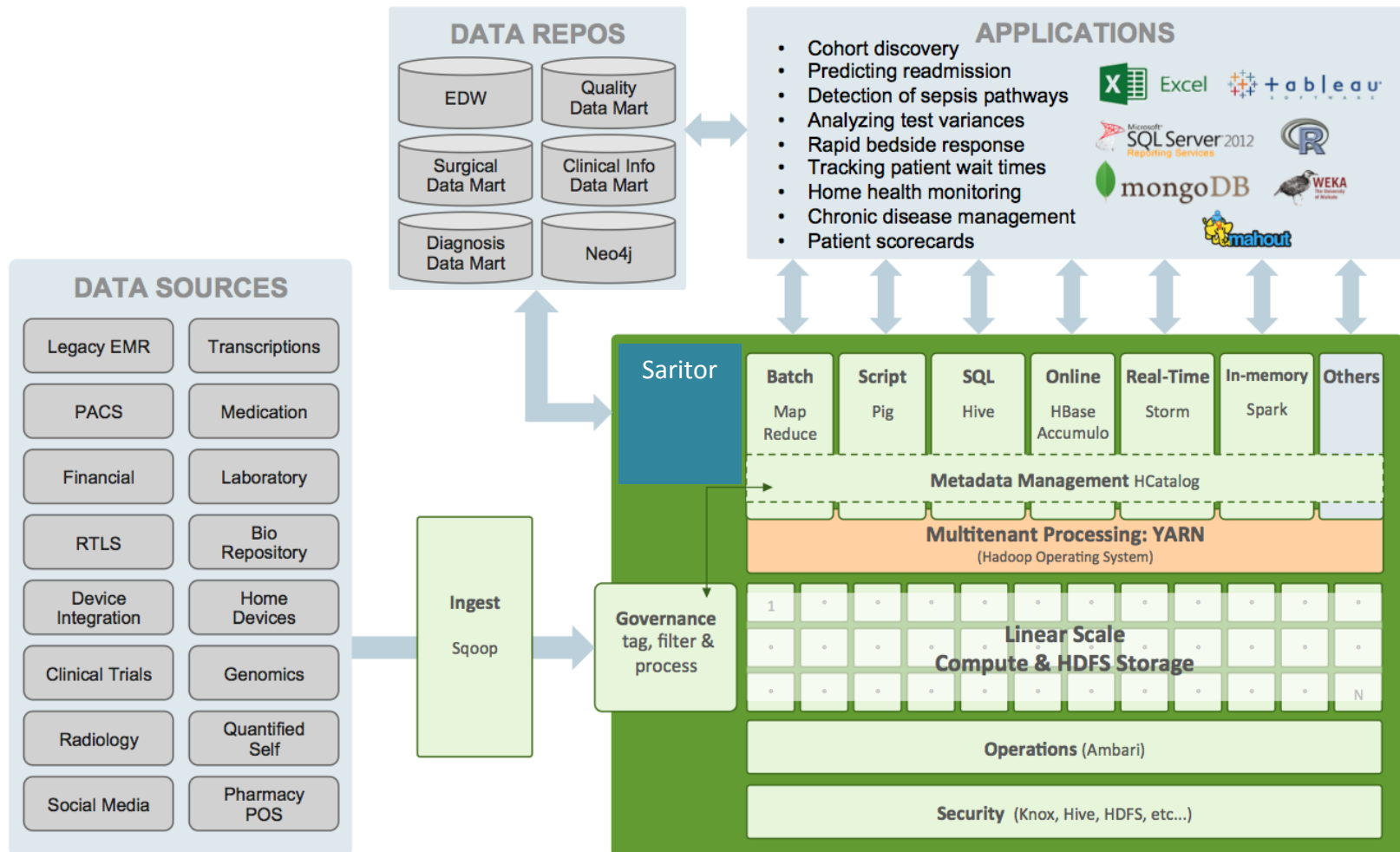
Contact Information

Charles Boicey

charles.boicey@stonybrookmedicine.edu

@N2InformaticsRN

Personalized Healthcare: From Population Data to Patient-Centered Outcomes

Nitesh Chawla, PhD

Frank Freimann Collegiate Chair of Engineering

Associate Professor of Computer Science and Engineering

Director, iCeNSA

Nitesh Chawla, PhD

UNIVERSITY OF NOTRE DAME

Time for Prospective Healthcare

Reactive Medicine → Prospective Medicine

Physician decision making is constrained by knowledge of complex disease factors and medical history.

Lab tests and family health history enhance physicians' assessments but generally focus only on a few diseases.

Medical intervention often begins only once a disease has emerged.

Health status of all patients are scored or categorized according to their risk to develop specific diseases.

Earliest onset of disease in patients are detected, health care needs predicted and appropriate preventive and chronic care services recommend.

Proactive personalized care plan for each individual is developed.

"Health care has been evolving away from a 'disease-centered model' and toward a 'patient-centered model.' In the older, disease-centered model, physicians make almost all treatment decisions based largely on clinical experience and data from various medical tests. In a patient-centered model, patients become active participants in their own care and receive services designed to focus on their individual needs and preferences, in addition to advice and counsel from health professionals." AHRQ.GOV

UNIVERSITY OF
NOTRE DAME

Two thousand years ago..

"It is far more important to know what person the disease has than what disease the person has," *Hippocrates*

Nitesh Chawla, PhD

UNIVERSITY OF
NOTRE DAME
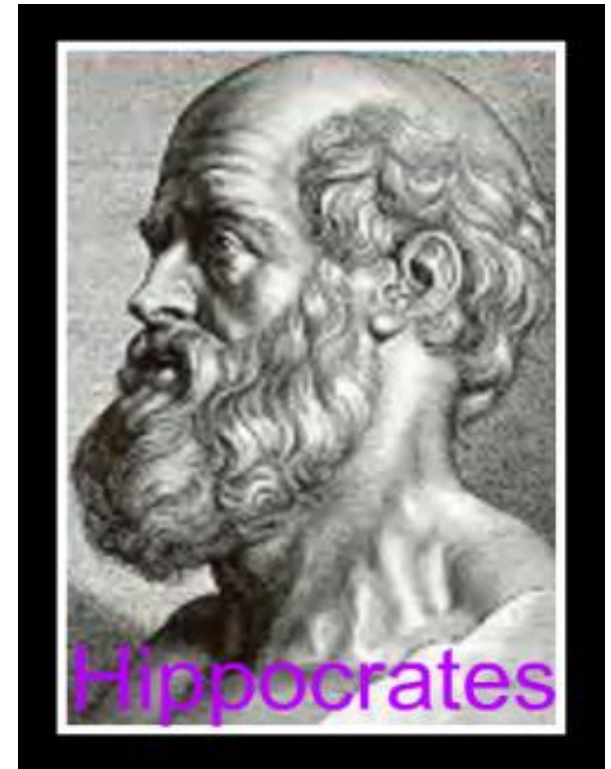
What are <u>my</u> disease risks? *A Personalized Approach*

"Determine <u>individual</u> risk of developing specific diseases, detect the disease's <u>earliest</u> onset, and <u>prevent</u> or intervene early enough to provide maximum benefit"

Nitesh Chawla, PhD

UNIVERSITY OF
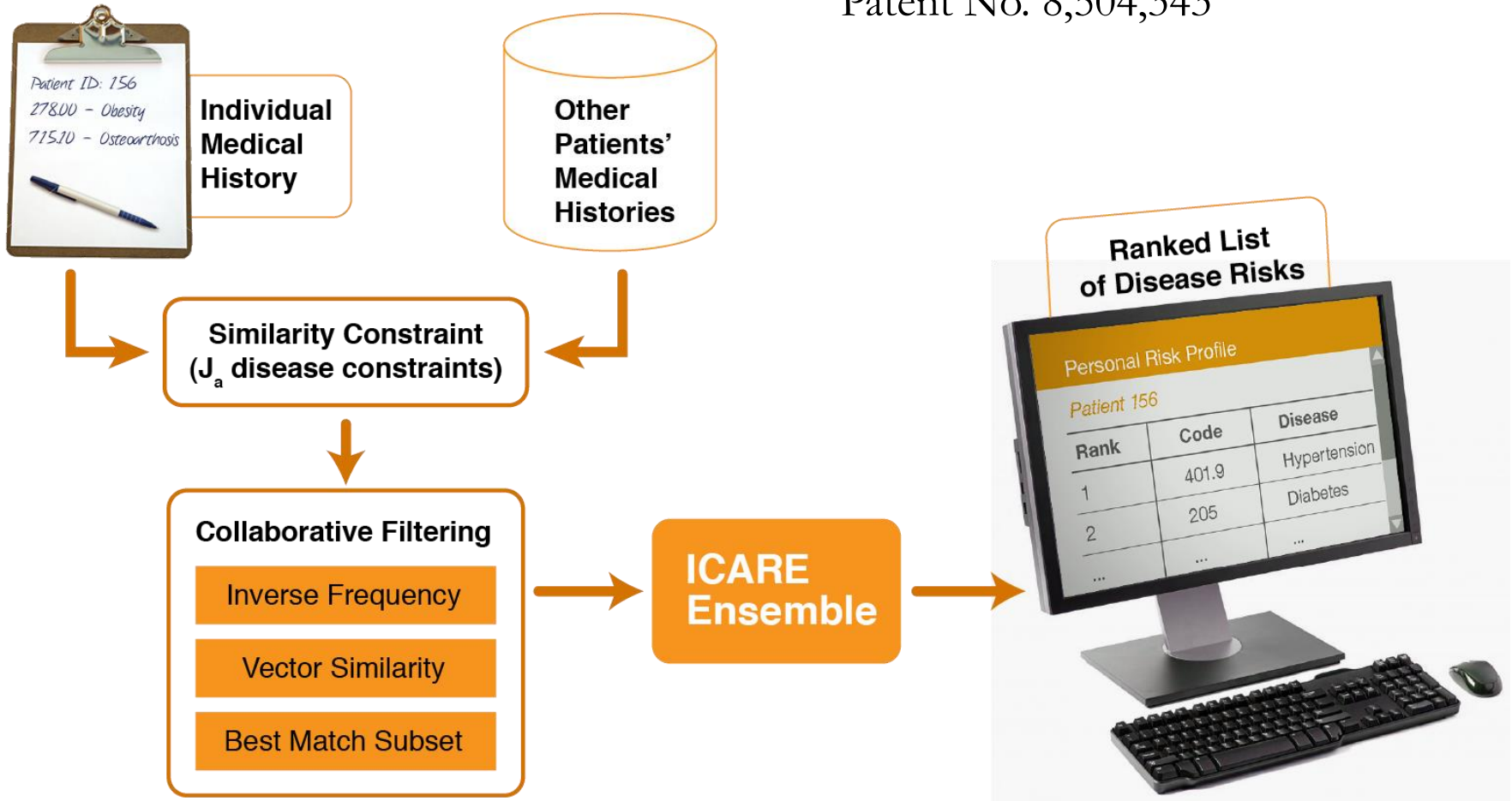NOTRE DAME

Similarities and shared experiences matter

CARE: Collaborative Assessment and Recommendation Engine

Patent No. 8,504,343

Empowering the patient and physician with the inferences drawn from millions of other patients

Nitesh Chawla, PhD

UNIVERSITY OF NOTRE DAME

CARE

Patent No. 8,504,343

Patient ID: 156
278.00 – Obesity
715.10 – Osteoarthosis

Individual Medical History

Other Patients' Medical Histories

Similarity Constraint (J_a disease constraints)

Collaborative Filtering

- Inverse Frequency
- Vector Similarity
- Best Match Subset

ICARE Ensemble

Ranked List of Disease Risks

Personal Risk Profile

Patient 156

Rank	Code	Disease
1	401.9	Hypertension
2	205	Diabetes
...	...	...

Nitesh Chawla, PhD

UNIVERSITY OF NOTRE DAME

"*Data is a vital raw material*"

Partnership is THE Enabler

- 13 Million ICD-9-CM data (Medicare)
  - In-patient
  - 32 Million visits over 4 years

ICD-9-CM Diagnosis 402
Hypertensive heart disease
  - 402 is a non-specific code that cannot be used to specify a diagnosis

ICD-9-CM Diagnosis 402.0
Malignant hypertensive heart disease
  - 402.0 is a non-specific code that cannot be used to specify a diagnosis

ICD-9-CM Diagnosis 402.00
Malignant hypertensive heart disease without heart failure
  - 402.00 is a specific code that can be used to specify a diagnosis
  - 402.00 contains 3 index entries

ICD-9-CM Diagnosis 402.01
Malignant hypertensive heart disease with heart failure
  - 402.01 is a specific code that can be used to specify a diagnosis
  - 402.01 contains 6 index entries

Nitesh Chawla, PhD

UNIVERSITY OF NOTRE DAME

Experimental Setup

- Predictions are only for future diseases

- Patient must have at least 5 visits

Round	Training	Testing
1	Visit 1	Visits 2-5
2	Visits 1-2	Visits 3-5
3	Visits 1-3	Visits 4-5
4	Visit 1-4	Visit 5

ICARE Results

	Baseline	3-digit ICARE
Top 20		
Coverage	0.321	0.513
Average Rank	7.326	5.668

Nitesh Chawla, PhD

UNIVERSITY OF
NOTRE DAME

Technology solve:
Empower, personalize and sustain

Initial Wellness Assessment → What are **my** risks?

Disease Management / **Nutrition/Exercise Program** ← What should **I** do?

Improved health and well-being ← Sustain: Wellness

Nitesh Chawla, PhD

UNIVERSITY OF NOTRE DAME
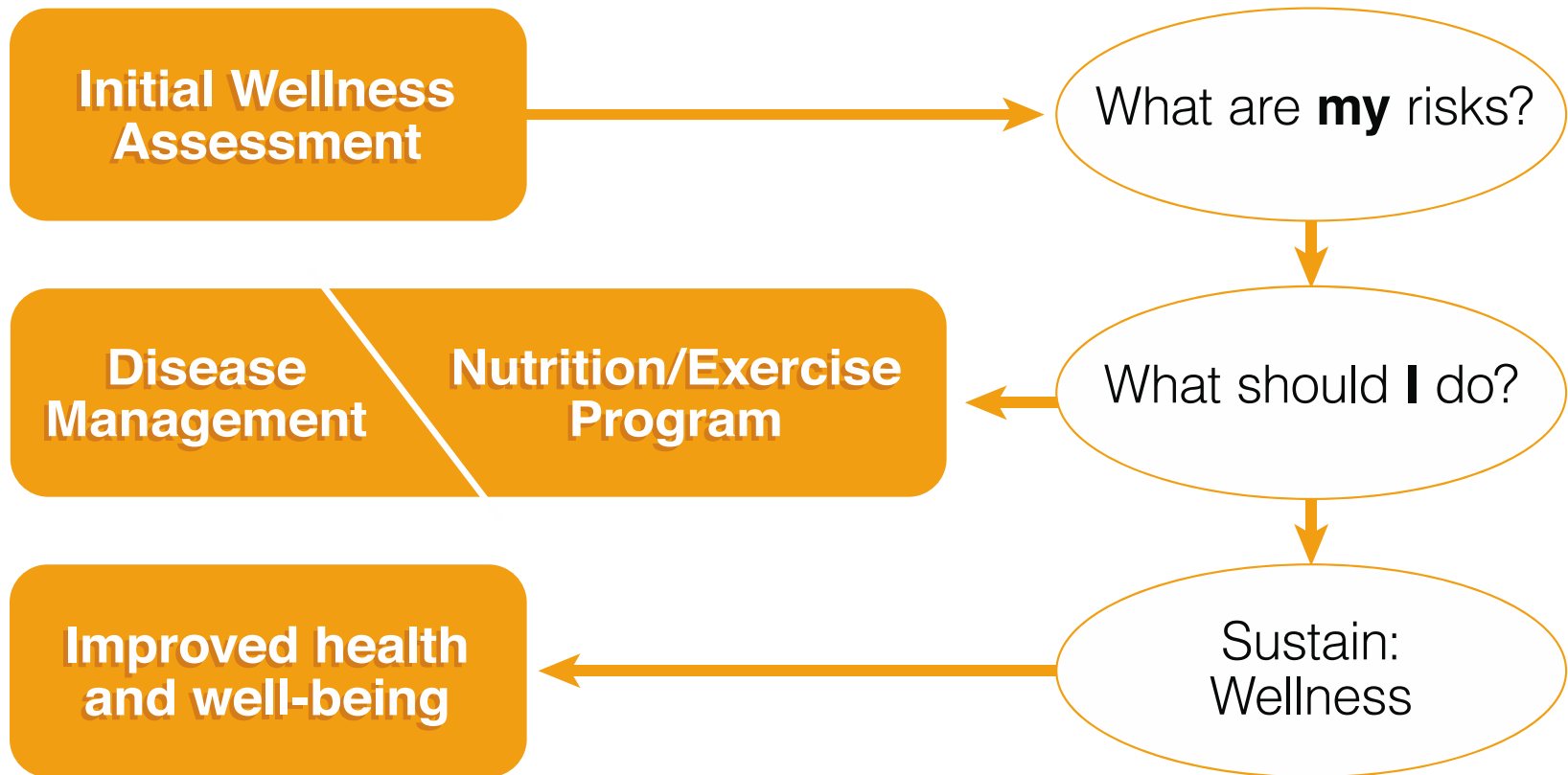
Me

Empower

Personalize

Connect

Sustain

eSeniorCare

Patient advocacy groups have way more power in adoption of health and wellness.

Nitesh Chawla, PhD

Thank you

Contact Info:

- nchawla@nd.edu

- http://www.nd.edu/~nchawla

- http://icensa.nd.edu

Nitesh Chawla, PhD

Advanced NLP for Electronic Health Records

Simon Beaulah, Director, Healthcare Strategy

simon.beaulah@linguamtics.com

Linguamatics

EHRs & Healthcare Challenges

Not for profit

Funders

Payers

ICT

Approvers

Prescribers

Education

Electronic Health Record

Medical devices

Government

Providers

Patient

Dispensers

Research

The challenge is to unlock the value of the huge investment being made in EHRs
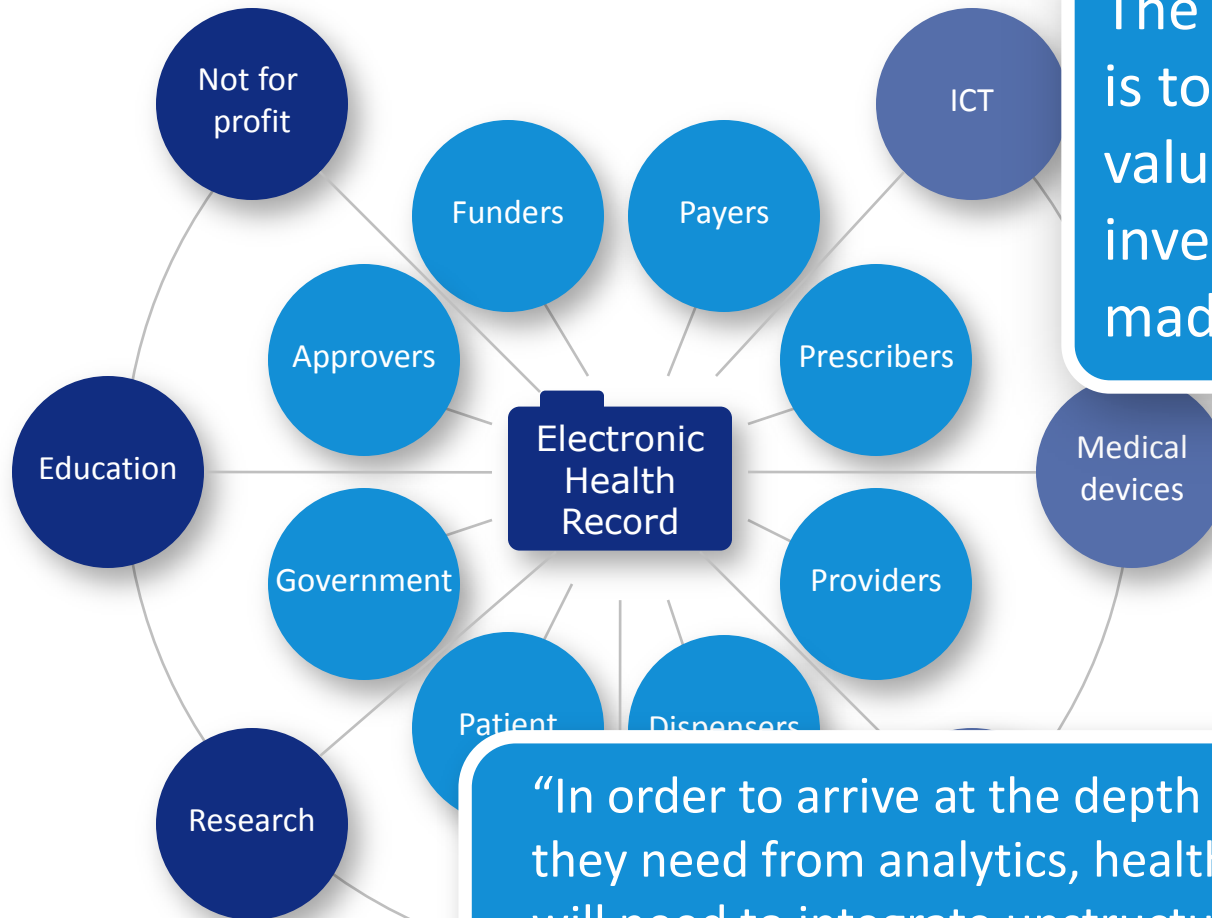
"In order to arrive at the depth of understanding they need from analytics, healthcare organizations will need to integrate unstructured data"

IDC Health Industry Insights

Linguamatics

Healthcare is in Transition

Healthcare needs to be a knowledge driven industry

- Enormous decision-making value in unstructured text if we can efficiently extract critical information from patient data

Vast and growing volumes of text

- Pathology, radiology and discharge reports not tractable with keyword search

Text mining/NLP transforms text into insights about patients

- Strong interest in Computer Aided Coding (CAC) but these systems are black box and only focussed on coding not information extraction.

- CAC can't cope with complex documents such as pathology and radiology

- Semantic normalization and enrichment essential

Challenges in Unstructured Patient Data

Different word, same meaning

cyclosporine

ciclosporin

Neoral

Sandimmune

Different expression, same meaning

Non-smoker

Does not smoke

Does not drink or smoke

Denies tobacco use

NLP

Different grammar, same meaning

5mg/kg of cyclosporine per day

5mg/kg per day of cyclosporine

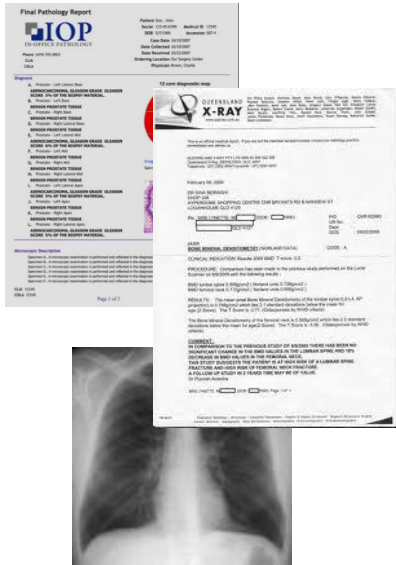cyclosporine 5mg/kg per day

Same word, different context

Diagnosed with diabetes

Family history of diabetes

No family history of diabetes

Linguamatics

NLP Transforms Text into Patient Insights

Turn text

Into structured data using sophisticated queries

To drive analytics

| Doc	Dimensions	First	Units	Second	Units	Third	Units	Hit
6280223	▶ Dimensions	2	mm	4	mm			1 On image 9 there is a 2 x 4 mm nodule within the right apex which is unchanged from prior examination which time it measured 3 mm.
6293739	Dimensions	1.9	cm					1 1.9 cm lesion in the right anterior abdominal wall at the level of the umbilicus, with central fluid density and enhancement in the periphery.
6362545	Dimensions	1.7	cm	0.9	cm	5.6	cm	▶ 2 There is a focal peripherally enhancing fluid collection along the medial aspect of the distal tibia measuring 1.7 x 0.9 x 5.6 cm.
5547811	Dimensions	2.6	cm	2.6	cm	2.7	cm	1 Intracranial extension is seen with a rim enhancing mass in the anterolateral left temporal lobe, measuring approximately 2.6 cm transverse x 2.6 cm AP x 2.7 cm craniocaudally.
6317842	▼ Dimensions	1.2	cm	1.2	cm			1 There are 2 high left parietal subcutaneous nodules that measure respectively 9 mm x 1.3 cm and 1.2 cm x 1.2 cm in largest dimensions as seen on image 11/248.
		9	mm	1.3	cm			1 There are 2 high left parietal subcutaneous nodules that measure respectively 9 mm x 1.3 cm and 1.2 cm x 1.2 cm in largest dimensions as seen on image 11/248.

Analytics

Enterprise Warehouse

Enterprise Biobank

Cancer Registry

Accurate results – only retrieves relevant results

Complete results – comprehensive and systematic

Linguamatics

NLP-Based Healthcare Use Cases
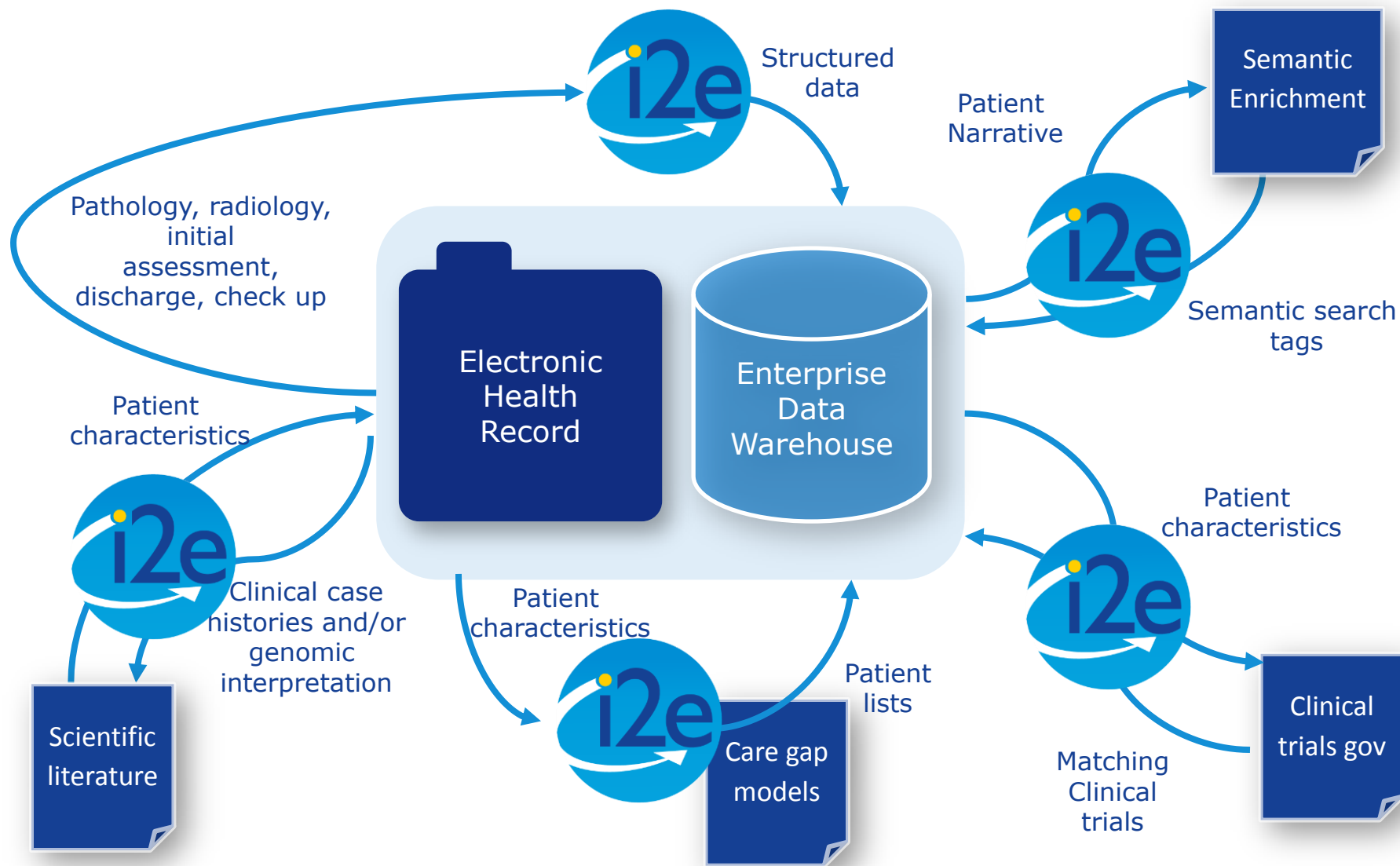
Linguamatics

KAISER PERMANENTE
PREDICTING PNEUMONIA FROM RADIOLOGY REPORTS

CHALLENGE

Diagnosis of pneumonia is a complex procedure requiring assessment of detailed radiologists' reports

SOLUTION

In collaboration with Linguamatics and I2E, Department of Research has constructed a model that predicts the presence or absence of pneumonia at 93% accuracy

BENEFIT

Large cohorts of patients can be assessed and specific cohorts selected based on complex patient documentation

GEORGETOWN UNIVERSITY
REALTIME DECISION SUPPORT USING iPADS

CHALLENGE

Published case histories provide valuable insights into disease comorbidity and treatments. Complex questions that cannot be easily answered, cause delays in treatment decisions.

SOLUTION

Georgetown University and Linguamatics have developed an application to enable rapid identification of case histories from PubMed during hospital rounds through iPad and Surface Tablets

BENEFIT

This rapid access to relevant data has saved hours and sometimes days of time and enabled faster decisions, leading to improved patient outcomes

www.linguamaticshealth.com

Predictive Model: Pulmonary Nodule Assessment

- The decision to investigate a pulmonary nodule with a biopsy is difficult due to the clinical risk of the procedure

- Predictive models rely on unstructured data

	Cancer Risk		
	Low	Intermediate	High
Nodule size, diameter (mm)	<8	8 to 20	>20
Age, yr	<45	45 to 60	>60
Prior cancer history	No prior cancer		Prior cancer history
Tobacco use (pack/day)	Never smoked	1	>1
Smoking cessation	Quit > 7 yr ago	Quit <7 yr ago	Never quit
Chronic obstructive lung disease	No COPD		COPD
Asbestos exposure	No exposure		Exposure
Nodule characteristics	Smooth	Lobulated	Spiculated

Linguamatics

GENOSPACE
MATCHING PATIENTS TO CLINICAL TRIALS

CHALLENGE

ClinicalTrials.gov's inclusion and exclusion criteria made matching patients to trials difficult to do automatically. Traditional NLP techniques were slow and not domain aware

SOLUTION

Genospace used I2E to automatically extract trial criteria in a structured form, including genetic needs, and load them into their database to support patient matching.

BENEFIT

Matches to trials are automatically made ensuring to the latest treatment options for patients.

Summary

- Application of analytics and NLP is key to future healthcare

- Complexity of human disease, associated specialties and social media means unstructured text is growing, not going away

- Use of NLP can impact patient care in numerous areas and be embedded into workflows

- Agile text mining provides a way to put it into practice now

- Contact me at simon.beaulah@linguamatics.com

www.linguamaticshealth.com

Linguamatics

General Discussion

- What are best practices for healthcare organizations to leverage EHR data in innovative ways?

- What barriers currently restrict the use of EHR data, and how can organizations overcome them?

  - How would you like to see EHRs improved to make them more amenable to secondary data use?

eHEALTH INITIATIVE
Real Solutions. Better Health.

Next steps

- Audio recording and slides will be available online at http://www.ehidc.org/issues/data-and-analytics/data-and-analytics-council-materials

- Next meeting: July 18, 2014

eHEALTH INITIATIVE
Real Solutions. Better Health.

Thank you!