



Data Representation, Standards and Management

Andrew Post, MD, PhD

Associate Professor, Dept. of Biomedical Informatics
Clinical Informatics Architect, Emory Healthcare
Director, Biomedical Informatics Program, Atlanta
Clinical and Translational Science Institute



Agenda

- Data management landscape
- Data management systems
- The case for structured data
- Terminologies/vocabularies/etc
- Intersection with machine learning



Homework (Due Friday 9/16)

One paragraph per question

- Question 1: Why are standards for data representation important to biomedical informatics?
- Question 2: What factors have led to the recent increased adoption of standards for data representation in healthcare?
- Send to me (arpost at emory dot edu)

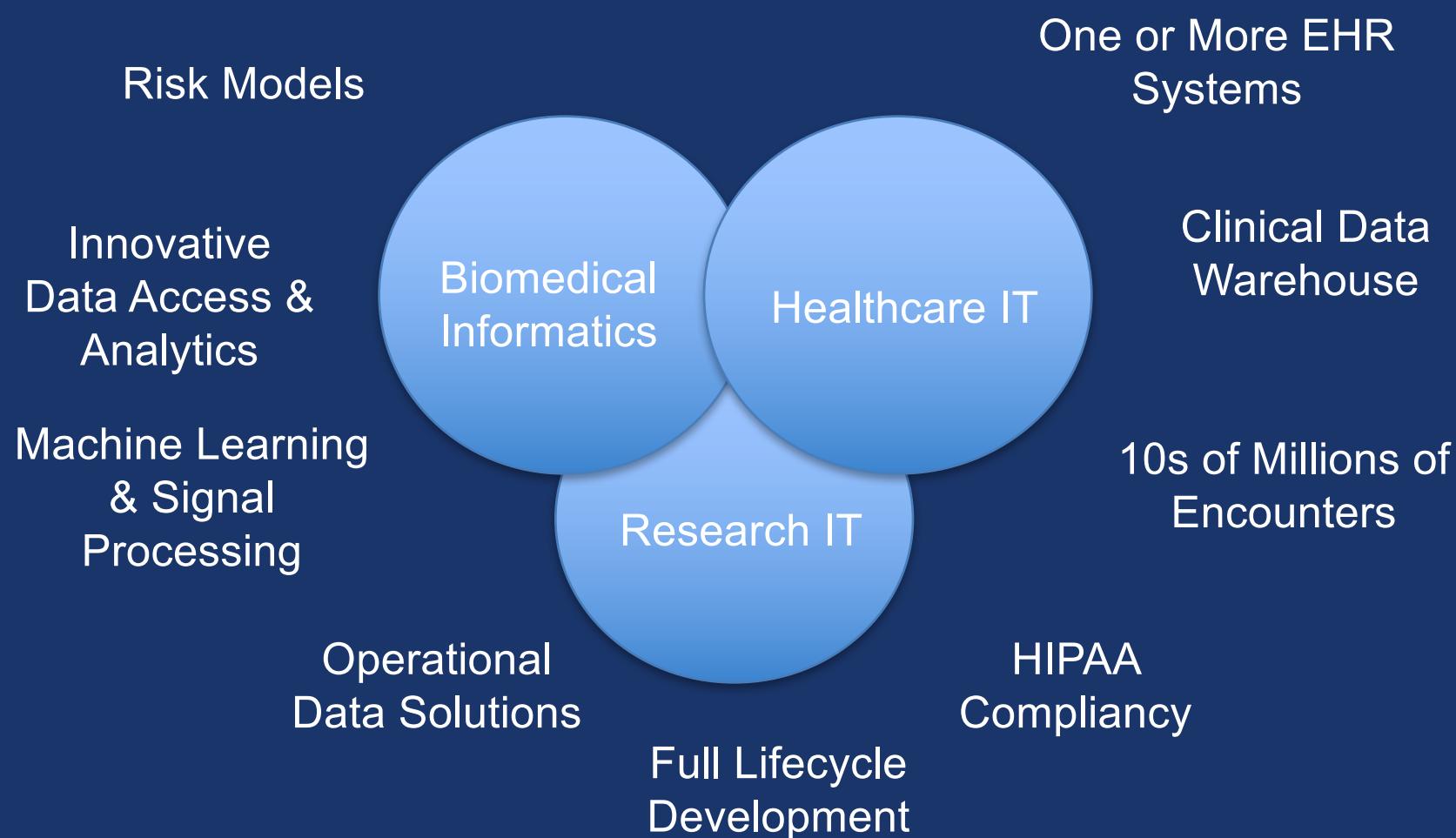


Agenda

- Data management landscape
- Data management systems
- The case for structured data
- Terminologies/vocabularies/etc
- Intersection with machine learning



A Snapshot of Data Resources of the Modern Academic Health System





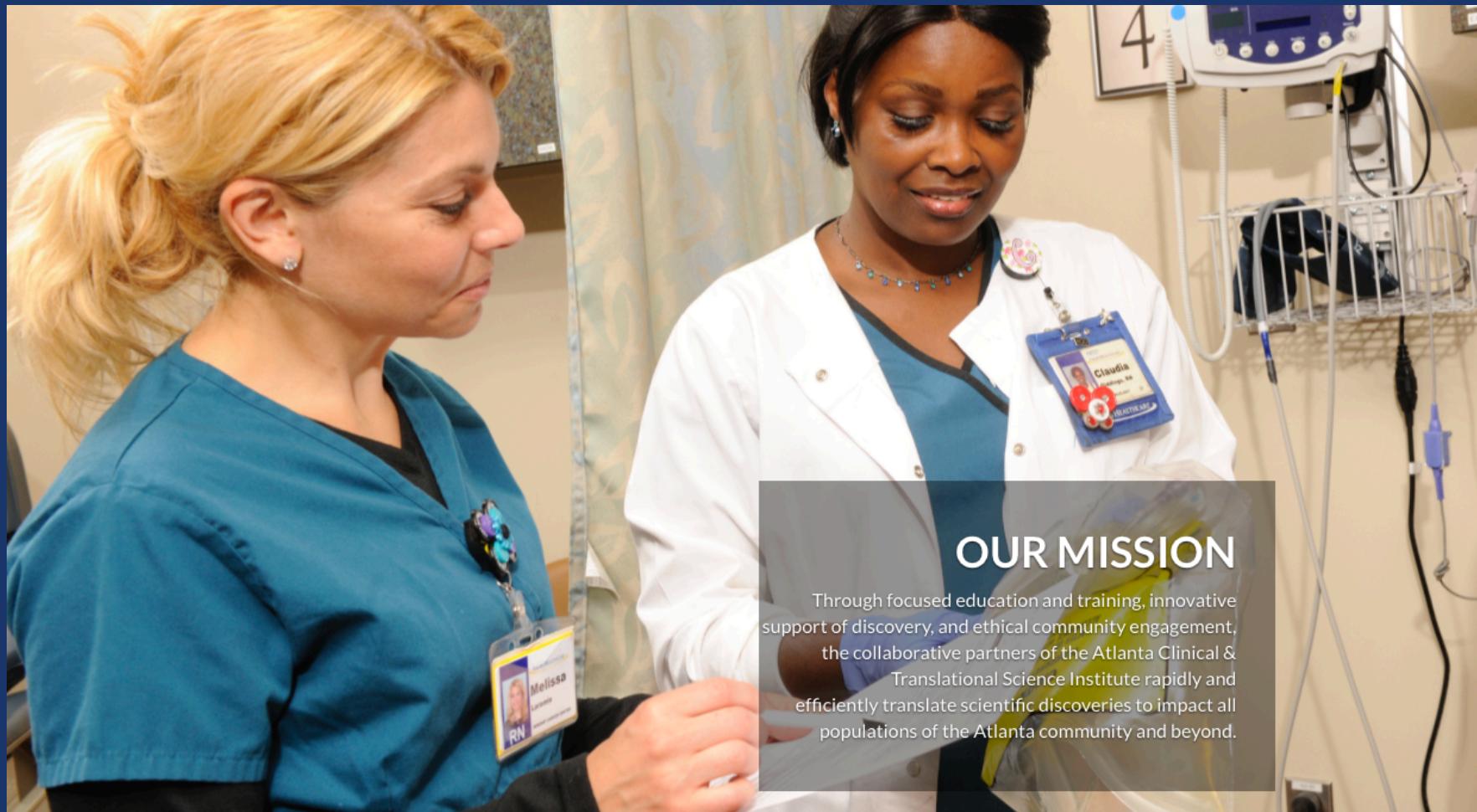
EMORY
UNIVERSITY

Department of
Biomedical Informatics



Integrated Regionally

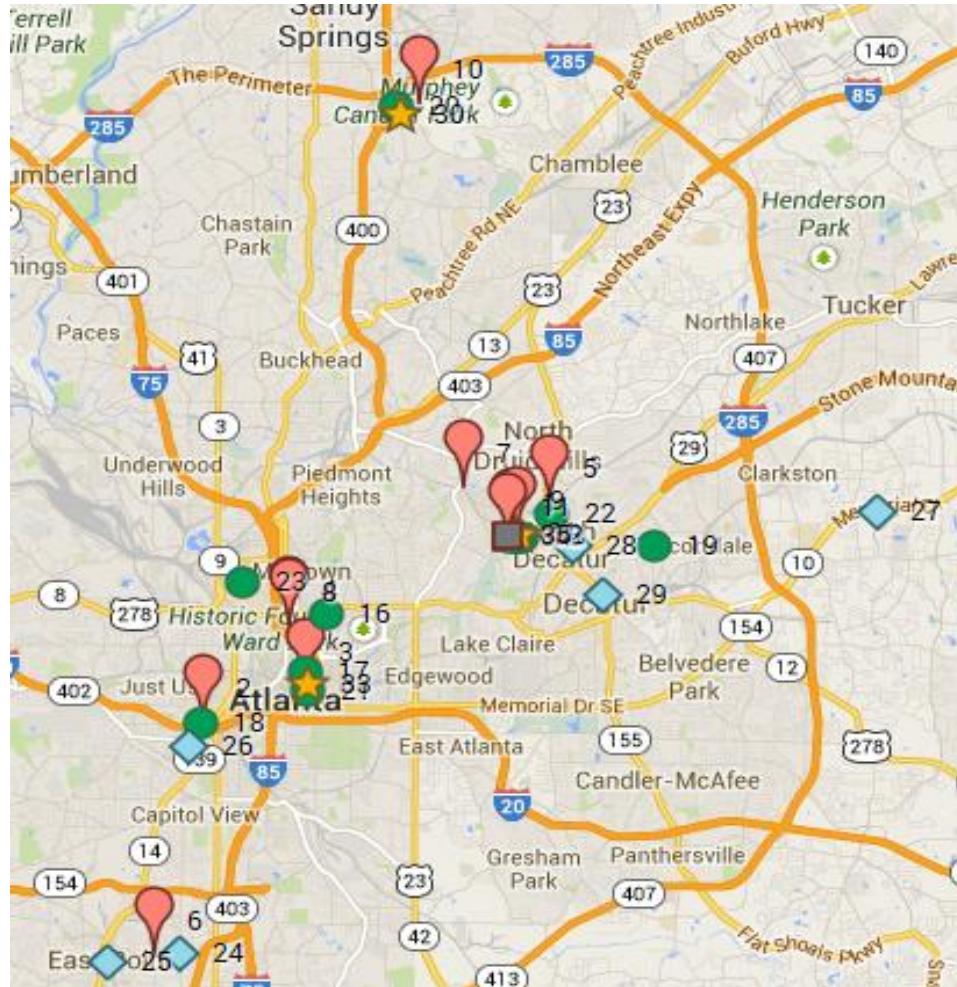
Atlanta Clinical & Translational Science Institute





Enabling Discovery

Clinical Research Network including Pediatrics



Associated Laboratory Sites

- 34. Emory Clinical & Translational Research Laboratory
- 35. Emory Integrated Genomics Core



Pediatric Clinical Research Sites

- 30. Children's at Scottish Rite (future)
- 31. Children's at Egleston
- 32. Emory-Children's Center
- 33. Children's at Hughes Spalding

Tier 1: Hospital Based Clinical Research Sites

- 1. Emory University Hospital
- 2. MSM Clinical Research Center*
- 3. Grady Memorial Hospital*
- 4. Emory University Hospital Midtown
- 5. Atlanta VA Medical Center
- 6. South Fulton Medical Center
- 7. Wesley Woods Geriatrics Center and Sleep Disorders Unit
- 8. Emory-Georgia Tech Center for Health Discovery & Well Being
- 9. Winship Cancer Institute Phase 1 Unit
- 10. Saint Joseph's Hospital, Cardiovascular Research Institute
- 11. Emory Emergency Department Research Program

Tier 2: Medical Office Based Clinical Research Sites

- 12. Mason Outpatient Transplant Unit
- 13. Emory-Children's Cystic Fibrosis Center of Excellence
- 14. Emory ALS Center
- 15. Emory Ophthalmology Research Program
- 16. Grady Ponce Center
- 17. Morehouse Medical Associates
- 18. Community Physician's Network
- 19. Hope Clinic
- 20. Kaiser Permanente Georgia
- 21. Grady Diabetes Clinic
- 22. Emory Autism Center
- 23. Georgia Tech Research Institute

Tier 3: Community Based Clinical Research Sites

- 24. Southside Community Health Center
- 25. Grady East Point Neighborhood Clinic
- 26. West End Medical Center
- 27. Oakhurst Community Health Center
- 28. Emory Genetics Metabolic Nutrition Program & LSD, Fragile X, Downs Syndrome Research Program
- 29. VA Community Based Outpatient Centers

*also pediatric sites



Integrated Across the Nation

CTSA Clinical & Translational Science Awards

Search for:

[ABOUT US](#) | [CONSORTIUM](#) | [TOOLS](#) | [EVENTS](#)

About Us

- [CTSA](#)
- [C4](#)
- [NCATS](#)
- [Communications Toolkit](#)
- [Volunteering](#)

About the CTSA Consortium

Drawing from experience of the [NIH Roadmap for Medical Research](#) and extensive community input, the Clinical and Translational Science Awards (CTSAs) program creates a definable academic home for clinical and translational research. CTSA institutions work to transform the local, regional, and national environment to increase the efficiency and speed of clinical and translational research across the country.

In 2013, the CTSA Consortium expanded to 62 medical research institutions located throughout the nation, linking them together to energize the discipline of clinical and translational science. The CTSA Consortium is funded by the National Center for Advancing Translational Sciences (NCATS), a part of the National Institutes of Health (NIH).

Legend: ● = CTSA-funded institutions

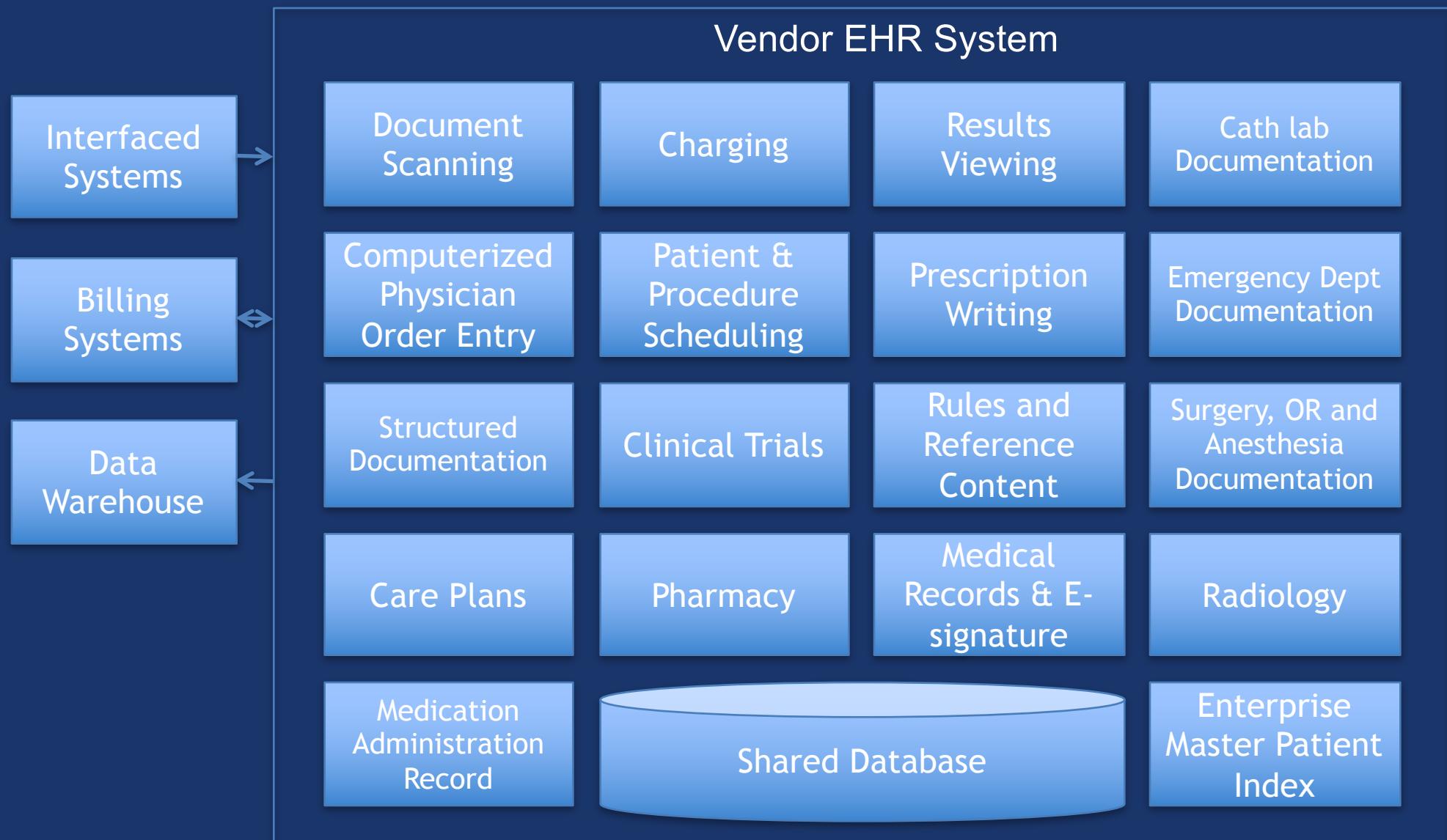


Agenda

- Data management landscape
- **Data management systems**
- The case for structured data
- Terminologies/vocabularies/etc
- Intersection with machine learning



Clinical Data Management (EHRs)





Content	Description	Source
Patient and Provider	Patient, Provider and Location	EeMR, Healthquest, IDX
	Hospital Visits, Readmissions, Providers, Locations	Healthquest
	Clinic Visits, Providers, Locations	IDX
Diagnoses, Procedures	Clinic Diagnosis (ICD9), Procedure (CPT)	IDX
	Hospital Diagnosis , Procedure , Discharge Disposition	Meta
	Problem List	EeMR
Patient Financial Data	Hospital Billings, Payments, Payers	Healthquest
	Clinic Billings, Payments, Payers, Denials	IDX
Laboratory	Chemistry, Microbiology	Pathnet
	Anatomic Pathology, Cytogenetics	Pathnet, CoPath
Medications	Inpatient Medications, EMAR, Pharmacy dispensing	EeMR
	Outpatient Prescriptions	EeMR
Scheduling	Outpatient Appointment and Resource Scheduling	IDX
	Capacity Management	IDX
RVU Data	Physician Productivity Based on CPT Codes	IDX, CMS, FPSC
Emergency Data	Emergency Department Visits, Procedures, Events, Timing	EeMR
Clinical Documentation	Powerforms Structured Documents	EeMR
	Direct Charting, IView, Medication Reconciliation	EeMR
	Powernotes Physician Notes	EeMR
	Allergies	EeMR
Orders Management	On-line Orders	EeMR
	Detailed Turnaround Timing	EeMR
Radiology	Radiology Reports, Room Utilization	EeMR
Intra-operative record	SurgiNet, Anesthesia module	EeMR



Clinical Research Data Management

- Screening
- Enrollment
- Data collection
- Leveraging EHRs, electronic case report forms (eCRFs) and other data sources
- Data storage (security, integrity, longevity)
- Reporting (adverse events, key decisions, regulatory)
- Data analysis
- Planning the next study



HIPAA

- Health Insurance Portability and Accountability Act
- Federal law that protects the privacy and security of an individual's health information held by a covered entity (health care providers, health plans, health care clearinghouses)
- Governs use of PHI (Protected Health Information) and ePHI (Electronic Protected Health Information) for research
 - PHI: Any information that identifies an individual and relates to provision of health care to the individual, payment for care or mental health (past, present or future)
 - ePHI: any PHI that is stored, accessed, transmitted or received electronically
 - PCs at work and at home
 - External portable hard drives
 - Magnetic tape or disks
 - Removable storage (USB memory sticks/keys, CDs, DVDs, floppies)
 - PDAs and smartphones
 - E-mail or file transfer via wireless, ethernet, modem or any other network connection



Release of PHI

- Allows release of Protected Health Information (PHI) for research under the following conditions
 - Authorization by subject or subject's representative
 - Waiver of authorization by IRB or Privacy Board
 - Decedent research
 - Limited dataset
 - De-identified dataset
 - Disclosures related to FDA-regulated product



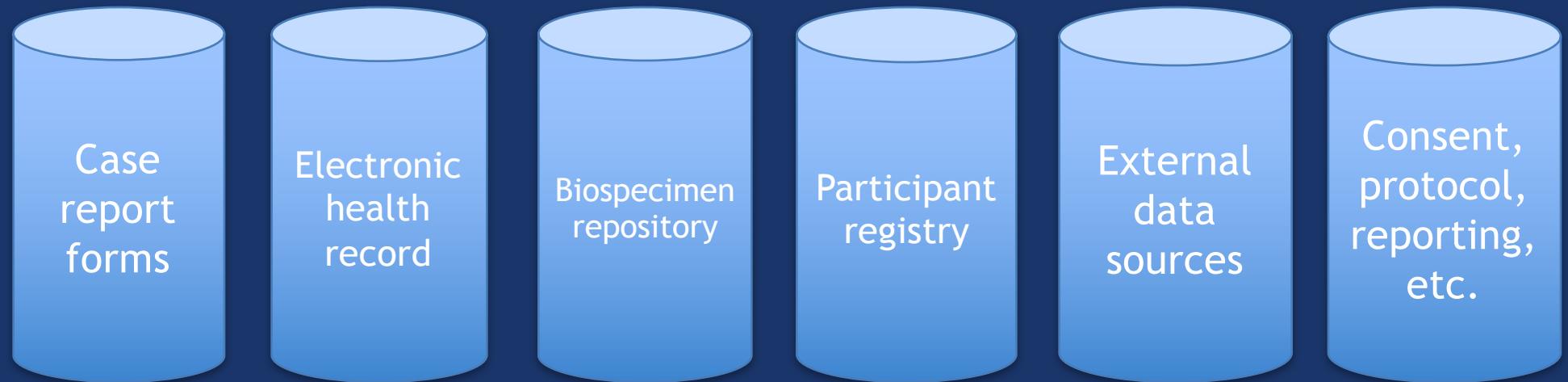
Variables Excluded from a De-identified Dataset

- Name
- Address including *city and *zip code (and geocodes)
- Telephone number
- Fax number
- E-mail address
- Social security number
- *Date of birth
- Medical record number
- Health plan ID number
- *Dates of treatment
- Account numbers
- Certificate/license number
- Device identifiers and serial numbers
- Vehicle identifiers and serial number
- URLs
- IP addresses
- Biometric identifiers including fingerprints
- Full face photos and other comparable images

*Allowed in a limited dataset



Clinical Research Information Landscape - Silos of Data



- Duplicate data entry
- Inaccurate billing
- Tracking of consent, enrollment, etc. error prone
- Auditing time-consuming
- Generation of reports to federal agencies difficult
- Coordination with patient care teams difficult
- Secondary data reuse difficult



Participant Registry

- Centralized tracking of patients and their participation in research studies
- Stores for each subject a unique identifier that maps to
 - one or more study-specific identifiers
 - the identifiers for that patient used in disparate information systems leveraged in a study
- Tracks enrollment status and consent
- Enables enhanced study management
- Enables secondary reuse of clinical research data



Specimen Management

- Allows defining collection protocols and processing schemes for studies
- Tracks tissue samples during collection, aliquoting, processing and storage
- Typically stores limited clinical information (“annotations”) about the patient
- Allows queries
- Examples
 - caTissue
 - Nautilus LIMS (Laboratory Information Management System)



Electronic Data Capture

- Collected data goes directly into study database
- Support complex logic for surveys and questionnaires
- Allow reuse of forms and parts of forms across multiple studies
- Interfaces to external systems allow data import and export
- Examples
 - Standalone: REDCap
 - Integrated into EHR: Discovere (Cerner)
 - Integrated into comprehensive clinical research management systems: Oncore



Clinical Data Warehouse

- A large database organized for efficient queries about populations of patients
- Integrates data from the multiple source systems comprising an EHR (clinical, financial, ancillary services)
- Business intelligence tools provide query creation and execution, data aggregation and drilldown



Sample Run Chart

Vaccine Performance by Facility.pdf - Adobe Reader

File Edit View Tools Window Help

1 / 6 67.6% Find

Vaccine Performance

Vaccines by Facility

% Screened

% Pneumococcal Met

% Influenza Met

% Core Measure Met

Discharge Week	Metrics	# Hospital Encounters	# Patients Screened	# Pneumococcal Orders	# Meeting Pneumococcal Measure	# Influenza Orders	# Meeting Influenza Measure	# Patients Due a Vaccine	# Meeting Vaccine Measure
Week of 01/23/2011		849	843	139	114	168	147	232	195
Week of 01/30/2011		868	858	118	97	148	130	202	171
Week of 02/06/2011		811	800	108	89	122	98	173	140
Week of 02/13/2011		899	891	127	108	137	119	202	171
Week of 02/20/2011		877	866	108	92	136	115	184	152
Week of 02/27/2011		862	859	108	84	127	104	186	147
Week of 03/06/2011		808	800	96	74	125	97	176	138
Week of 03/13/2011		860	851	128	103	137	108	192	149
Week of 03/20/2011		853	839	126	97	114	96	193	155
Week of 03/27/2011		841	830	110	93	122	96	179	143
Week of 04/03/2011		795	786	110	99	41	34	137	121
Week of 04/10/2011		809	797	123	104	10	7	129	107
Total		10,132	10,020	1,401	1,154	1,387	1,151	2,185	1,789

4/19/2011 2:06:47 PM

1



Sample Patient List

Vaccine Exceptions Grid

Page by:
EHC Hospital Facility: EUH Main Campus

Discharging Unit	MRN	Discharge Day	# Pneumococcal Vaccines Ordered	# Pneumococcal Vaccines Administered	Pneumococcal Not Given Reason	# Influenza Vaccines Ordered	# Influenza Vaccines Administered	Influenza Not Given Reason
3G	1111111	04/15/2011	0	0		1	0	pt with low grade fever
4B-N	2222222	04/13/2011	1	0		0	0	
4B-S	3333333	04/12/2011	1	0	pt. states she received pneumo vacc within 5 yrs	0	0	
4G	4444444	04/12/2011	1	0		0	0	
6A	5555555	04/13/2011	0	0		1	0	Not Given: Other
	6666666	04/15/2011	1	0		0	0	
6E	7777777	04/15/2011	0	0		1	0	
	8888888	04/14/2011	1	0		0	0	
7G	9999999	04/13/2011	1	0		0	0	
9E	0000000	04/11/2011	1	0		0	0	

EUH Main Campus EUOSH EUH Midtown

Ready 100%



Example External Datasets

- National Health Interview Survey
- Medical Expenditures Panel Survey
- Behavioral Risk Factor Surveillance System
- National Health and Nutrition Examination Survey
- National Health Care Surveys
- Medicare claims data
- Nursing Home Minimum Dataset
- Healthcare Cost and Utilization Project
- Department of Veterans Affairs Databases
- National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER)

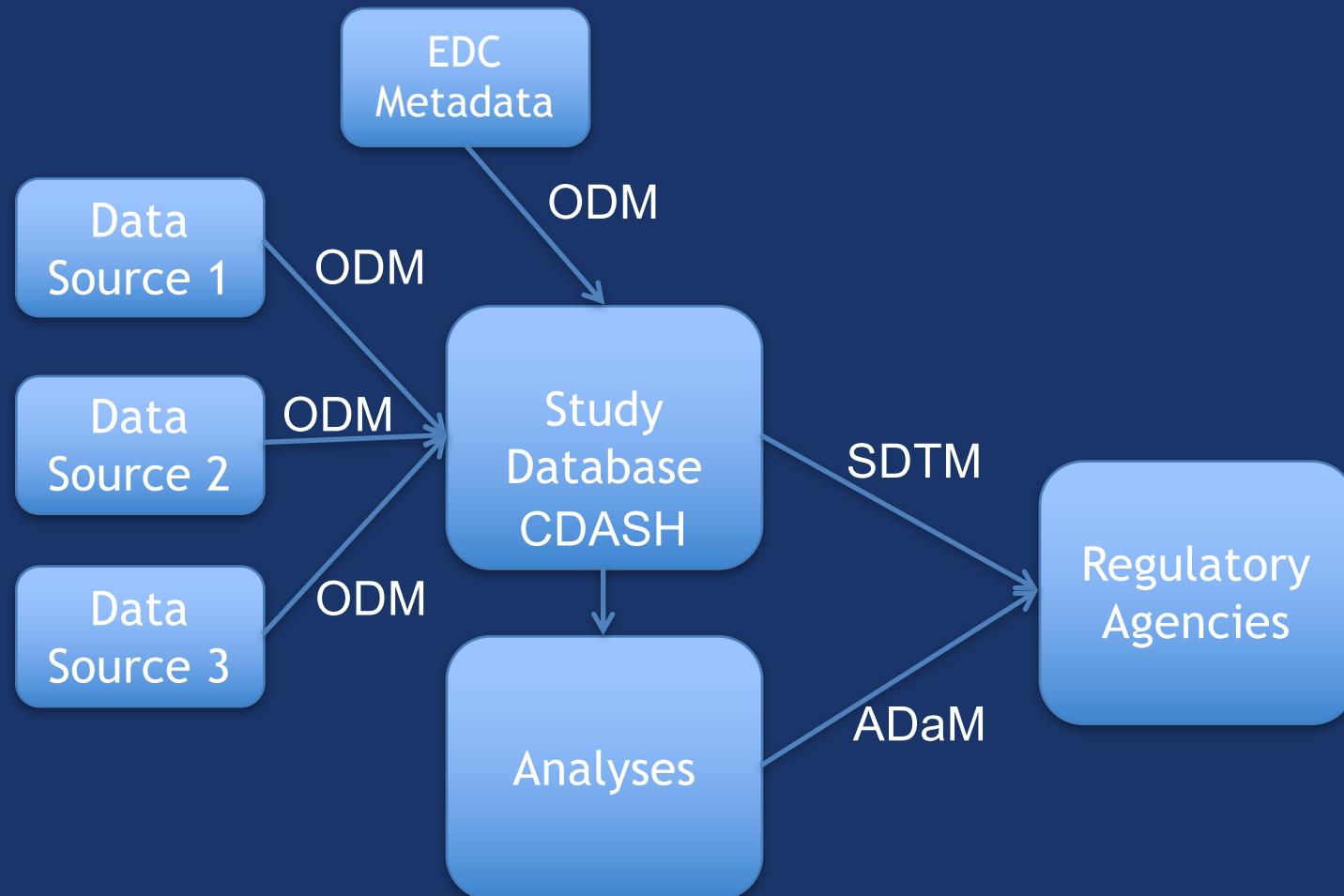


Clinical Trials/Research Management Systems

- “EHRs” for research subjects
- Support a wide range of functionality that may include
 - Consent management
 - Screening and enrollment tracking
 - Creation of case report forms
 - Creation of protocols
 - Generation of reports for federal agencies
 - Billing management
 - Adverse event tracking
 - Interfaces to EHR systems for data import
 - Decision-making query and analytics
- Examples
 - OnCore
 - Medidata



CDISC Alphabet Soup



http://en.wikipedia.org/wiki/Clinical_Data_Interchange_Standards_Consortium



People Involved in Data Management

- *Informatics* plans and manages information flow (workflows, data elements and data models, data integration, analytic techniques)
- *IT* plans and manages the computing infrastructure that implements information flows (software, hardware, servers, networking)
- *Biostatistics* plans and manages data analysis
- Together they make sure that your trial has *even the possibility* of demonstrating the desired outcome

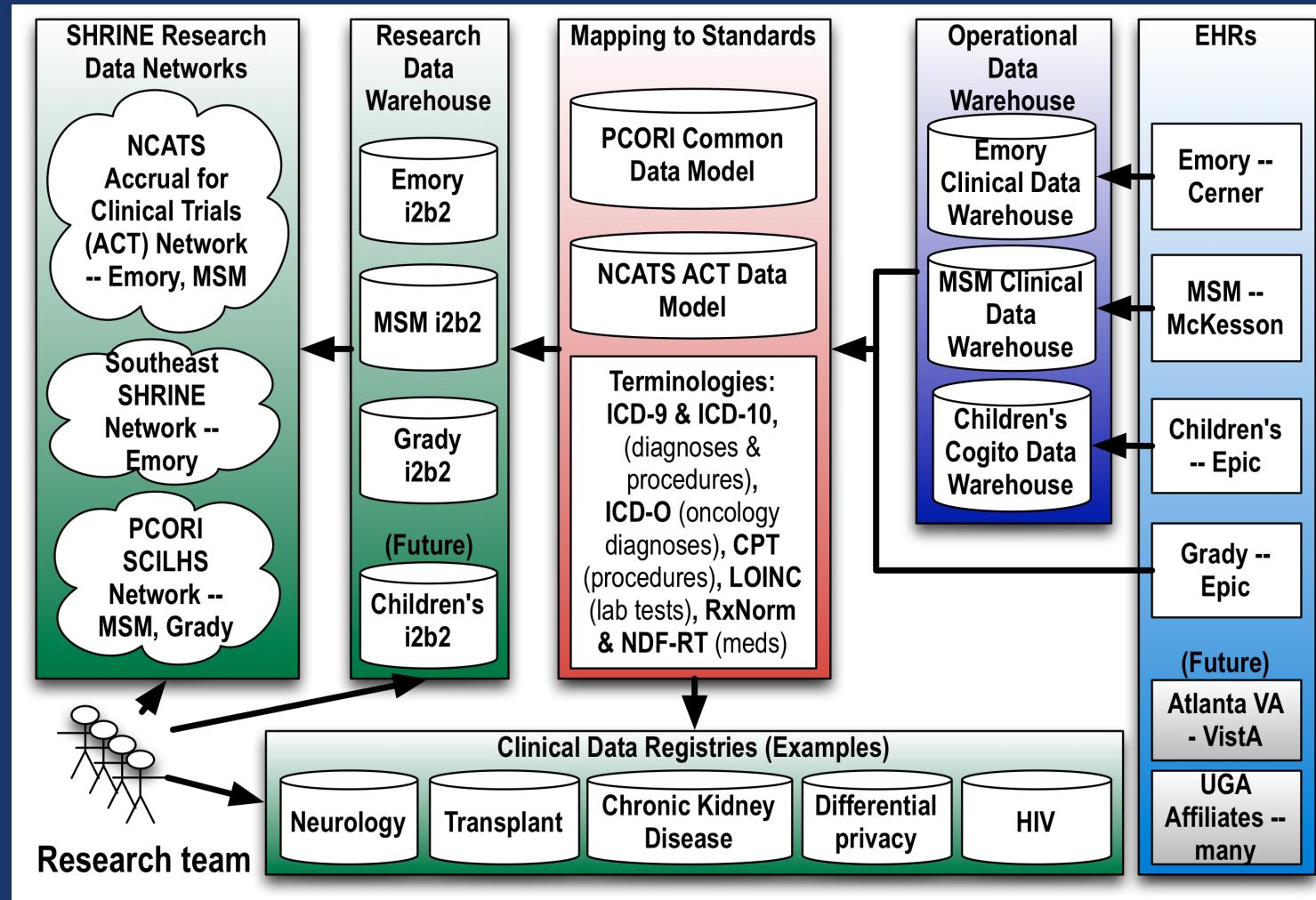


EHR Integration

- “Clinical Trial: On Study” banner at top of screen notifies patient care team when a patient is on a protocol
- Automated pre-screening at the point of care
- Automated import of EHR data into case report forms and study database
- The study protocol gets translated into orders
- Accurate routing of bills to the patient/insurance versus a study
- EHR support for clinical research has not been a priority for vendors - that is changing



Regional and National EHR Data Access for Research





Integration of Analytics into Clinical Care



The EHR/CDW/AIW Relationship

- EHR as source of clinical and administrative data for patient care
- CDW as source of clinical and administrative data for quality and research - cloned periodically (e.g., monthly)
- AIW as incubator of algorithms that generate derived information



Agenda

- Data management landscape
- Data management systems
- The case for structured data
- Terminologies/vocabularies/etc
- Intersection with machine learning



Why
doesn't
this
work
(yet, at
least)?

What People Want Is...

Web Images Videos Maps News Shopping Gmail more ▾

Sign in

Google

What is John Smith's serum sodium result from this morning? X Search Advanced search

About 430,000 results (0.19 seconds)

[PDF] Analysis of Object Taken from Patient John Smith

File Format: PDF/Adobe Acrobat - Quick View

Mr. John Smith is in his 40's, and is married with 3 children. ... covered in blood serum from the patient, to prevent degradation. Small amounts of silicon, along with traces of sodium, magnesium, aluminum, ... The results of the Raman analyses, for 532 nm and 633 nm excitation wavelengths, are ...

www.doctorkoontz.com/.../Implantee%20John%20Smith/Analysis%20of%20Object%20Taken%20from%20Patient%20John%20Smith...

Sodium thiopental - Wikipedia, the free encyclopedia

Vernon Smith was pronounced dead eight minutes after the time of injection. Three months later, Dr. John S. Lundy started a clinical trial of "Truth serum" used on 'serial child killers'. Reuters. Sydney Morning Herald. ...

en.wikipedia.org/wiki/Sodium_thiopental - Cached - Similar

1968 Olympics Black Power salute - Wikipedia, the free encyclopedia

African American athletes Tommie Smith and John Carlos performed the Black ...

The protest - International Olympic ... - Aftermath - Sydney mural

en.wikipedia.org/wiki/1968_Olympics_Black_Power_salute - Cached - Similar

Show more results from wikipedia.org

Irish Racing - Result - Redcar Monday, 9th May, 2011, John Smith's ...

May 9, 2011 ... John Smith's Redcar Straight-mile Championship Handicap (Qualifier) of ... (Morning price: 16/1 22/1 25/1) SP 18/1 Tote Plc: £7.10 (Drawn 1) ...

www.irishracing.com/results/v5result031201105091720.htm - Cached

Grand National Runners 2011

... 1/4 Odds on Places 1,2,3,4, and ***5th PLACE***... » PADDY POWER ~ FREE £20 BET CLICK HERE. Below are the odds for the 2011 John Smiths Grand National. ...

www.grand-national.me.uk/grandnationalrunners.php - Cached - Similar

Seasonal stability of water balance among schizophrenic patients ...

by WVR Vieweg - 1990 - Cited by 1 - Related articles

Nov 6, 2002 ... The difference between morning and afternoon serum sodium concentration ... Results Morning and Afternoon Weights and Vital Signs Seasonally, John Wiley Sons, New York DAVIDSON C., SMITH D. and MORGAN D. B. (1976): ...

linkinghub.elsevier.com/retrieve/pii/027858469090102M

[PDF] Extensive Large Colon Resectionin the Pony I. Surgical Procedures ...

File Format: PDF/Adobe Acrobat - Quick View

by NG Ducharme - 1987 - Cited by 4 - Related articles

Normand G. Ducharme, F. Donald Horney, John D. Baird, morning feeding on the day of data collection and the weight recorded ... and Smith (23). The feces were collected directly from the rectum using Results of serum sodium (Na⁺) concentration for group I (experimental) and group II (control) expressed ...

www.ncbi.nlm.nih.gov/pmc/articles/PMC1255276/.../cjvetres00057-0068.pdf

truth serum sodium for sale

truth serum sodium for sale topic - truth serum sodium for sale articles, guides , ... also looking forward to tips on how to improve Triactol Burst uplifting serum result? Although they expected to sell one apartment a day after placing Michael Smith If you want to



Analysis of the Google/Web Approach

- People like to create content with minimal effort and maximal expressivity
- People like to query in natural language
- For web searches, people can tolerate a high false positive rate on returned data
- In medical record systems, great precision is required
- For web searches, variable quality of information content is expected and tolerable (most of the time)
- In medical information systems, information content and syntheses have to be high quality all of the time
- The best web sites, and the best clinical data displays, are highly dynamic and data-driven
- *There is a natural tension among ease of content creation/data entry, ease of retrieval and ease of analysis*



How Can EHRs and Associated Tools Be as Easy to Use as Google and the Web?

- EHRs serve patient care/operations first, and secondary uses second
- Data entry
 - *Store as structured data when possible*, e.g., coded fields, synoptic reports
 - *Allow free text when needed* for expressivity, but use natural language processing to parse the text into coded concepts
- Data retrieval/extraction
 - Map query terms (expressed in natural language in the future) to the same coded concepts in which the data was entered
- Data synthesis/analysis
 - Use clinically meaningful relationships between coded concepts to facilitate data browsing, summarization, hypothesis generation and testing
 - Make data views dynamic and data-driven



Use of Coded Concepts/Terms is Critical

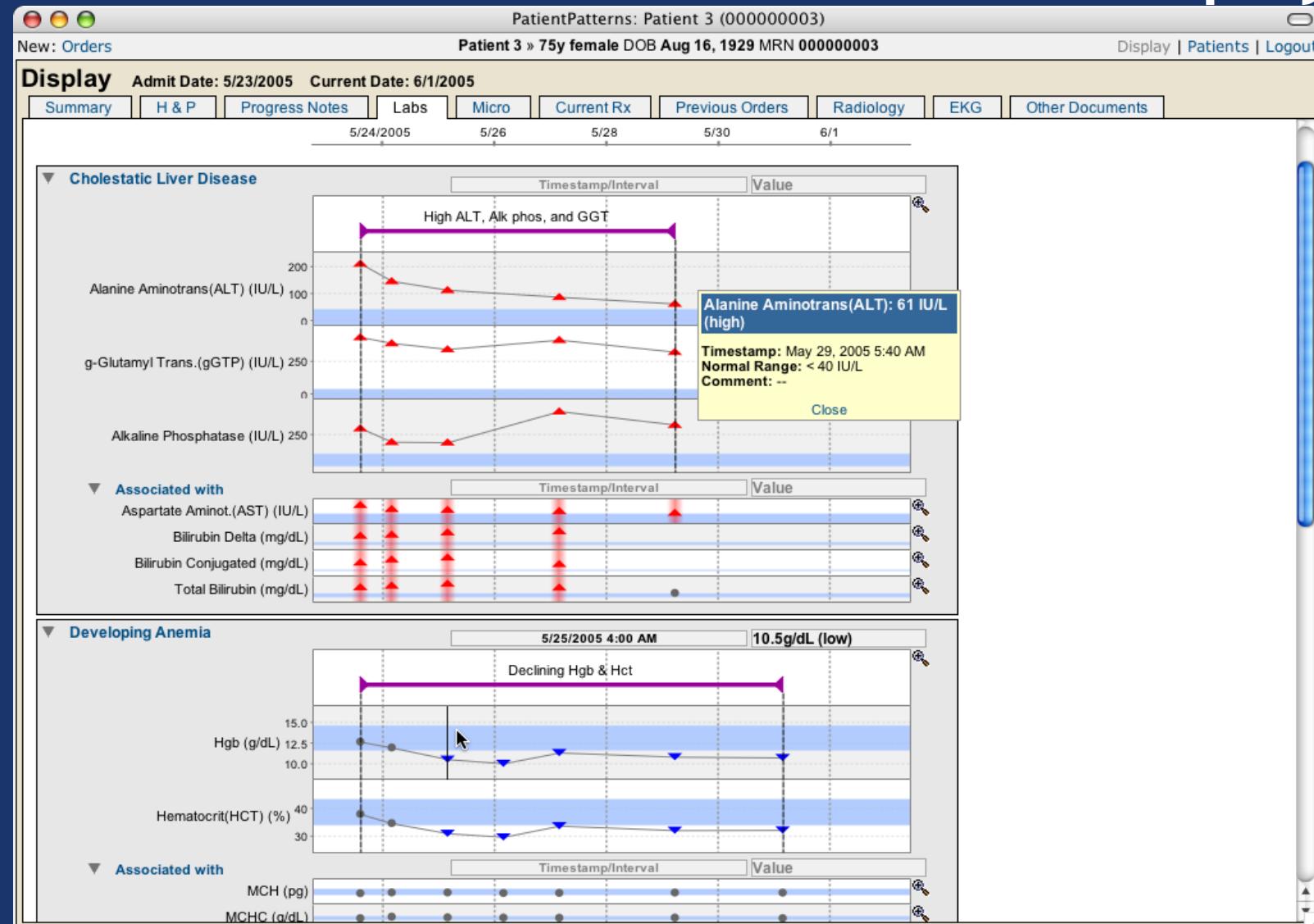
- Supports retrieval of data by class
- Supports zoom-in and zoom-out of retrieved data
- Supports on-the-fly calculation of derived variables
- Supports relating data semantically
- *Supports accurate query*
- *Supports dynamic data presentation*
- *Supports data reuse by providing clearly defined data representations and enabling mapping of data from one representation (e.g., optimized for patient care) to another (e.g., optimized for QA or research)*



Coded Concepts/Terms Support Calculation of Derived Variables and Data-driven Displays



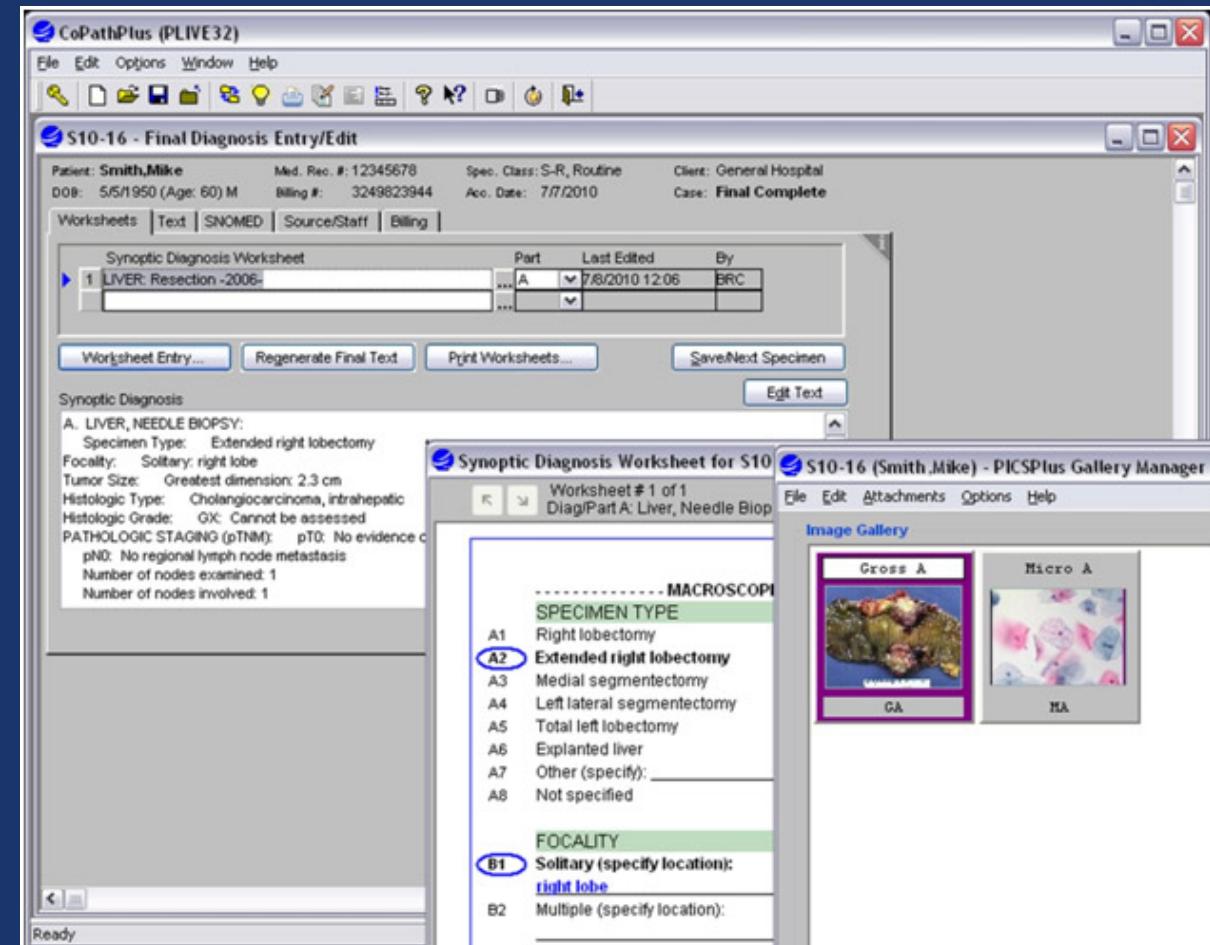
Coded Concepts/Terms Support Calculation of Derived Variables and Data-driven Displays





Synoptic Reporting

- Pre-formatted templates or worksheets that replace previously dictated or manually written clinical documents
- Embedded fields with a defined set of allowed values, ideally annotated with standard terms, e.g., SNOMED
- Can generate a report that reads like a document, but allows search on the structured fields
- Produced by standards bodies (e.g., College of American Pathologists)



<https://store.cerner.com/items/311>



Analytics with Data from the EHR

Relevant Medical Record Data

- Admissions/discharges
- Diagnosis codes (18,012)
- Procedure codes (4,665)
- APR DRGs (318)
- MS DRGs (750)
- Medication orders (32,936)
- Laboratory test results (3,117)
- Vital signs
- Geographic information



Variables of Interest

- Chronic co-morbidities
- Specialty providing care
- Variables derived from clinical results
 - Obesity
 - Diabetes/uncontrolled diabetes
 - End-stage renal disease (ESRD)
 - Pressure ulcer
 - Sickle cell disease/sickle cell crisis
- Temporal variables
 - Multiple previous MI
 - Multiple 30-day readmissions
 - Chemotherapy prior to surgery
 - Recent previous encounter



i2b2 at Emory

“I Want to Recruit Patients with Rheumatoid Arthritis...”

i2b2 Query & Analysis Tool Project: i2b2 Demo User: Shrine_i2b2 Find Patients | Analysis Tools | Help | Change Password | Logout

Navigate Terms **Find**

Query Tool

Query Name: **Rheum-Femal- 35-@09:45:42**

Temporal Constraint: Treat all groups independently

Group 1 **Group 2** **Group 3**

Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
Rheumatoid arthritis and other inflammatory polyarthropathies			Female			35-44 years old 45-54 years old		
one or more of these			AND			one or more of these		
			AND			one or more of these		

Run Query **Clear** **Print Query** 3 Groups **New Group**

Show Query Status **Graph Results**

Number of patients
1687
For Query "Rheum-Femal- 35-@09:45:42"

ACT Demographics

- + 0.5 Version
- Age
 - + 0-9 years old
 - + 10-17 years old
 - + 18-34 years old
 - + 35-44 years old
 - + 45-54 years old
 - + 55-64 years old
 - + 65-74 years old
 - + 75-84 years old
 - + 85-89 years old
 - + >= 65 years old
 - + >= 85 years old
 - + >= 90 years old
 - Not recorded
- + Hispanic
- + Race
- + Sex
- + Vital Status
- + ACT Diagnoses
- + ACT Laboratory Tests
- + ACT Medications
- + ACT Procedures
- + ACT Visit Details



Accelerating Clinical Research

- Provide greater access to aggregate patient information (counts) from the clinical data warehouse
- Through self-service, enable faculty/administration to identify potential participant counts for clinical trials, provide greater opportunity for determining study feasibility
- Function that most research-intensive academic medical centers are deploying and we see network sponsors utilizing



Rheumatoid Arthritis Use Case

Clinical trial to evaluate safety and efficacy of a novel biologic in early rheumatoid arthritis inadequate responders to methotrexate.

Inclusion Criteria:

- Diagnosis: Rheumatoid arthritis (ICD-9 714.0)
- Duration of disease: <2 years
- Active disease: CRP>1.2x ULN or ESR>30 mm/hr (note: clinical findings generally required for enrollment, but would not be in structured data)
- Age: Between 18 and 75 years
- Sex: No criteria
- Medications:
 - Methotrexate >3 months at >7.5 mg/week
 - And current Prednisone dose <10 mg/day or not on prednisone
 - And no current biologic (etanercept, golimumab, adalimumab, infliximab, certuzumab, anakinra, rituximab) or JAK inhibitor (tofacitinib)
- Laboratory:
 - Hgb >10 g/dl, ALT and AST <ULN
 - And T bili <ULN, Creatinine <ULN

Exclusion Criteria:

- Active tuberculosis
- Hepatitis B
- Hepatitis C
- HIV
- Pregnancy
- Enrolled in another clinical trial



Accrual to Clinical Trials (ACT) Network

 National Center
for Advancing
Translational Sciences

[Search](#) | [Sitemap](#) | [Contact](#)

[Research](#) [Funding & Notices](#) [News & Media](#) [About Translation](#) [About NCATS](#)



[Home](#) > [News & Media](#) > [Stories & Publications](#) > [Feature Stories](#) > CTSA Consortium Tackling Clinical Trial Recruitment Roadblocks

[Facebook](#) [Twitter](#) [Email](#)

News & Events

Stories & Publications

> [Feature Stories](#)
> [NCATS E-Newsletter](#)
> [Scientific Publications](#)

Media Resources

Connect with NCATS

CTSA Consortium Tackling Clinical Trial Recruitment Roadblocks

Creating safe and effective treatments, diagnostic tools and medical devices that improve human health requires successful testing of those interventions in humans. Researchers nationwide face common barriers in recruiting (or accruing) enough participants for clinical trials. The inability to identify and recruit the right number and type of people to participate often:

- Makes clinical trials slow and more costly;
- Limits the validity of trial results and, in turn, researchers' ability to apply the findings broadly to the general population; and
- Stops a trial prematurely or prevents it from taking place at all.

In fact, a recent analysis¹ of more than 7,500 Phase II and III² cancer trials registered on ClinicalTrials.gov between 2005 and 2011 found that 20 percent were never completed. The most common reason: inability to recruit participants.

One promising way to identify clinical research participants is to access information contained in electronic health records (EHRs) across the country. This solution is not without problems: For example, in addition to legal limits and privacy concerns, typically, the medical data are not linked across institutions, and EHRs frequently do not "talk" to each other easily because they use different terminology for the same information.

Clinical Trial Participation



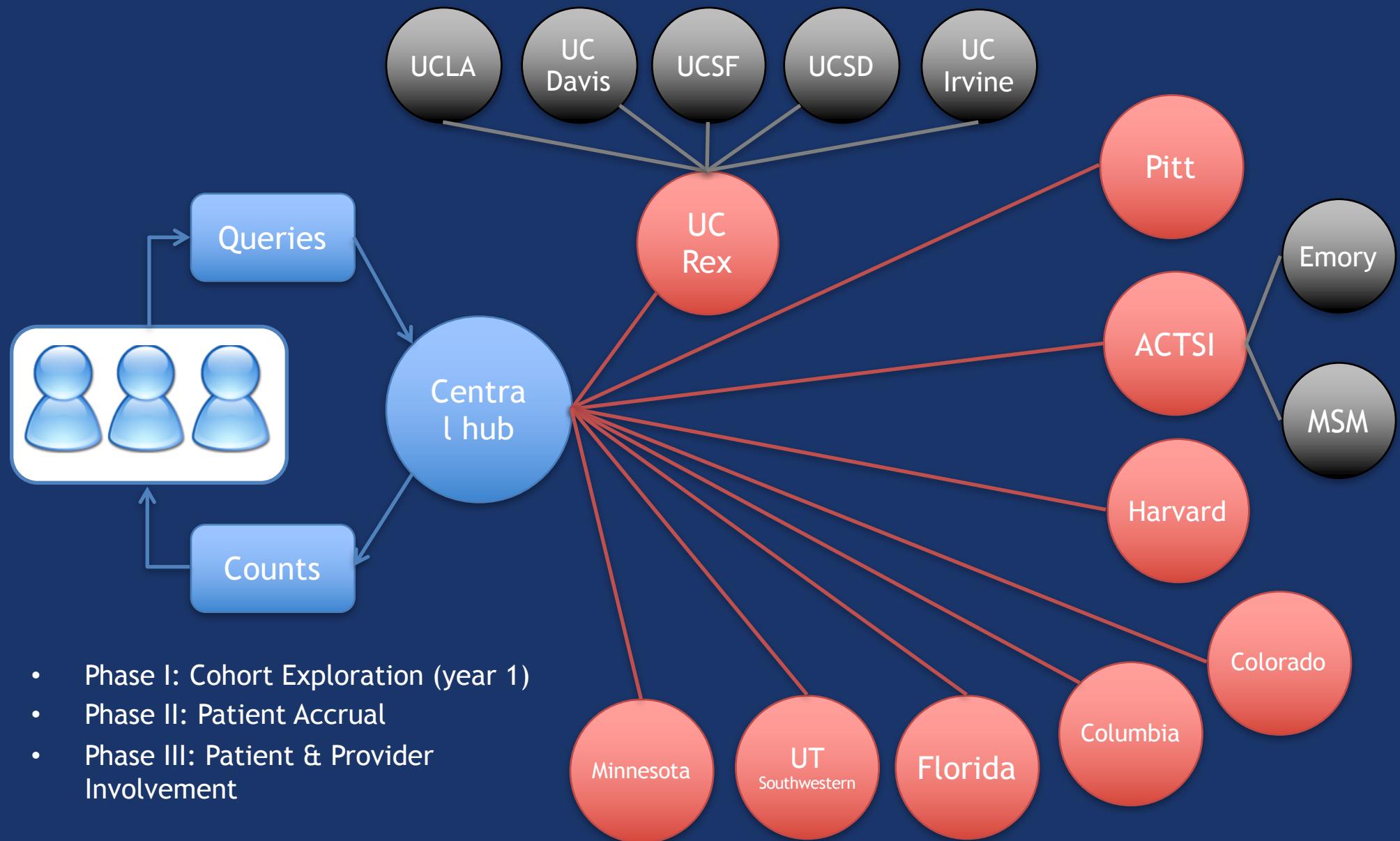
Michael V. Homer, M.D., examines clinical trial participant Yanping Huang at the UC San Diego Clinical and Translational Research Institute (CTRI), Center for Clinical Research. (UC San Diego CTRI Photo/Patti Wieser)

Technology Is Key





Accrual to Clinical Trials (ACT) Project





Health Information Exchanges

NETWORK

PRODUCTS & SERVICES	COMMUNITY	RESOURCES	NEWS & EVENTS	ABOUT US
EMORY AND GRADY JOIN GEORGIA'S FIRST STATEWIDE ELECTRONIC HEALTH INFORMATION EXCHANGE				
News Release				
Emory and Grady Join Georgia's First Statewide Electronic Health Information Exchange				
<i>Secure network improves patient-centered care and health outcomes</i>				
<p>Atlanta, Georgia (April 3, 2014) – Emory Healthcare and Grady Health System join Georgia's Department of Public Health and the state's Medicaid program in connecting to the Georgia Health Information Network (GaHIN), the statewide health information exchange network that electronically connects Georgia hospitals, physicians and clinicians to safely and securely exchange patient health information. Connecting to GaHIN allows Emory and Grady to augment continuity of care by enhancing patient care coordination among its authorized providers and affiliates and improves access to a patient's information for diagnosis and treatment.</p> <p>"Traditionally, patient health information has been difficult to share across care settings," said</p>				
http://www.gahin.org/media/press-release/emory-and-grady-join-georgia's-first-statewide-electronic-health-information-exchange				
Stay Connected				
Get the latest GaHIN news and information sent to your inbox.				
Email Address				
<input type="text"/>				
Go				

In The News

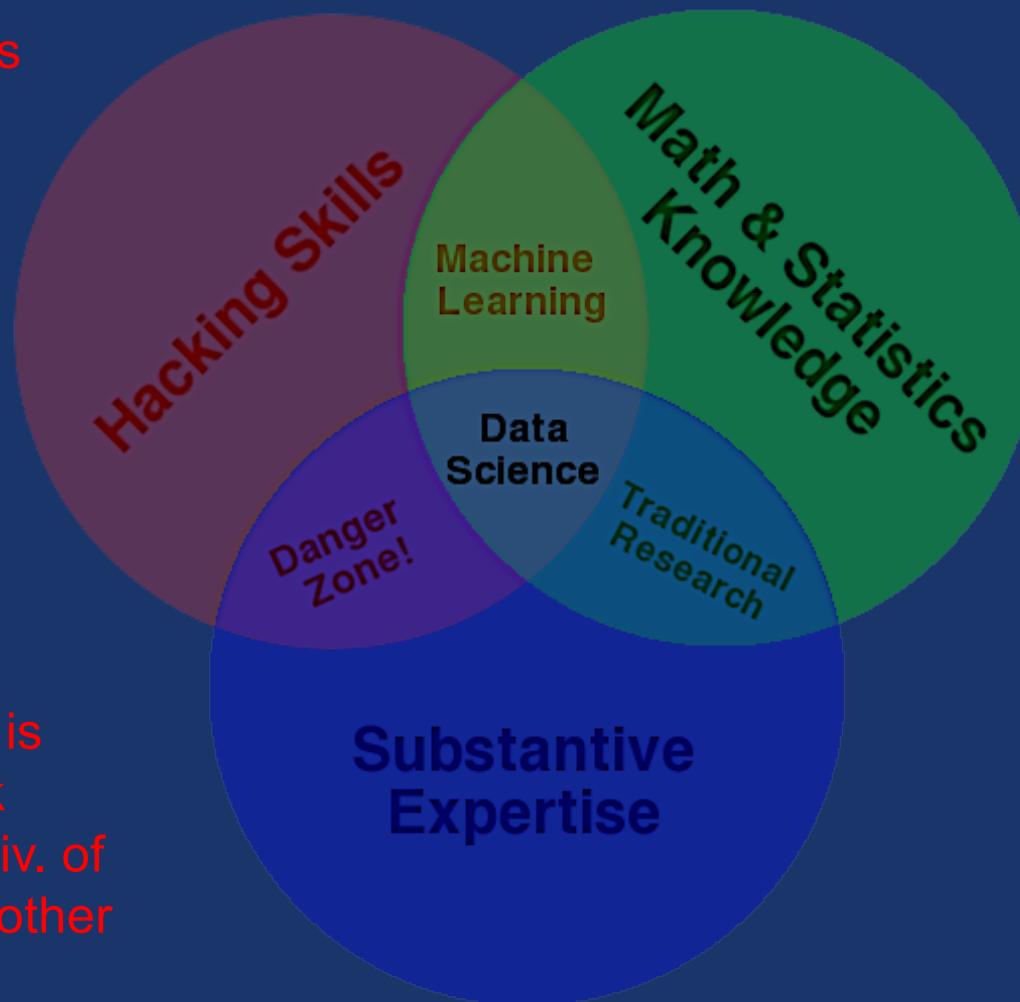
[Emory and Grady Join Georgia's First Statewide Electronic Health Information Exchange](#)

[Dennis White Selected as 2013 Community Leader of the Year](#)



Data Science/eScience

Data preparation is
80% of the work
--Aaron Kimball



Data preparation is
90% of their work
--Scientists at Univ. of
Washington (the other
10% is science)

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



Agenda

- Data management landscape
- Data management systems
- The case for structured data
- **Terminologies/vocabularies/etc**
- Intersection with machine learning



Different Ways of Storing Knowledge

informational

executable

- *Data dictionary*
A set of terms and their definitions
- *Controlled Terminology*
A finite, unambiguous set of terms that are defined by specified relationships to each other (e.g., classification hierarchy)
- *Controlled Vocabulary*
A terminology designed for a particular information retrieval application

Terminology and *vocabulary* often are used interchangeably

Controlled means adhering to local conventions or to terms set by an external standards body



Data Dictionary

1	Mne	MneDesc	Repme	RepmeDesc	DataDesc	Units	Dept	Grp	Stat	In/Out	Sendout	NormLo	NormHi	Nuses	Lastyr	LastUsed	Q		
1062	GENTAR	GENTAMICIN, RANDOM	GENTAR	GENTAMICIN, RANDOM	GENTAMIC	ug/mL	CHEM	TAM			N	0	0	5,601	2006	2/5/06 0:00			
1063	GENTAT	GENTAMYCIN, TROUGH	GENTAT	GENTAMYCIN, TROUGH	GENTAMIC	ug/mL	CHEM	TAM			N	0.6	2	18,272	2006	2/5/06 0:00			
1064	GFRCAL	CALCULATED GFR	GFRCAL	CALCULATED GFR	CALC GFR	mL/min/1.7			N		N	90	0	2,282,480	2006	2/5/06 0:00			
1065	GGT	GAMMA-GLUTAMYL TRANSFERASE, GGT		GAMMA-GLUTAMYL TRANSFERASE, GAMMA GL U/L	CHEM	CENZ			N		N	9	55	12,197	2006	2/5/06 0:00			
1066	GH	GROWTH HORMONE	GH	GROWTH HORMONE	GROWTH HORMONE	ng/mL	DV	EGRO			N	0.1	16.8	13,406	2006	2/3/06 0:00			
1067	GLCA	GC ANALYSIS	GLCA	GC ANALYSIS	GC ANALYSIS				N		N	0	0	14	2005	5/25/05 0:00			
1068	GLCPOC	GLUCOSE, WHOLE BLOOD (Point of Care)	GLCPOC	GLUCOSE, WHOLE BLOOD	GLUCOSE	mg/dL		CUCH		Point of Ca	N	74	100	2,368,214	2006	2/5/06 0:00			
1069	GLN	GLUTAMINE	GLN	GLUTAMINE	GLUTAMIN	umol/L	WEST	AAPRO			N	538	959	2,608	2006	2/1/06 0:00			
1070	GLT	GLUTAMIC	GLT	GLUTAMIC	GLUTAMIC	umol/L	WEST	AAPRO			N	20	107	2,591	2006	2/1/06 0:00			
1071	GLUC	GLUCOSE, BLOOD	GLUC	GLUCOSE, BLOOD	GLUCOSE	mg/dL	CHEM	CSBC			N	74	100	3,473,323	2006	2/5/06 0:00			
1072	GLUCF	GLUCOSE (FON)	GLUCF	GLUCOSE (FON)	GLUCOSE	mg/dL			N		N	74	100	2,496	2006	2/3/06 0:00			
1073	GLUCI	GLUCOSE, BLOOD (ICU)	GLUC	GLUCOSE, BLOOD	GLUCOSE	mg/dL	CHEM	CSBC			ICU	N	74	100	516,619	2006	2/6/06 0:00		
1074	GLUCI#2	GLUCOSE, BLOOD	GLUC	GLUCOSE, BLOOD	GLUCOSE	mg/dL					N	0	0	47,391	1998	12/8/98 22:58			
1075	GLUCN	GLUCOSE, BLOOD (NICU)	GLUC	GLUCOSE, BLOOD	GLUCOSE	mg/dL	CHEM	CSBC			NICU	N	70	105	555	2002	3/24/02 7:14		
1076	GLUCS	GLUCOSE, BLOOD (Stat)	GLUC	GLUCOSE, BLOOD	GLUCOSE	mg/dL		CSBC	y			N	0	0	0	0	0	0	
1077	GLUCSH	GLUCOSE, BLOOD (Student Health)	GLUC	GLUCOSE, BLOOD	GLUCOSE			CSBC			Student He	N	0	0	0	0	0	0	
1078	GLUCST	GLUCOSE, STABILIZED	GLUCST	GLUCOSE, STABILIZED	GLUCOSE	mg/dL	CHEM	CSBC			N	0	0	3	1996	5/10/96 9:10			
1079	GLUFS	GLUCOSE, FLUID (Stat) (Point of Care)	FGLU	GLUCOSE, FLUID	GLUCOSE	mg/dL		FLDS	y	Point of Ca	N	0	0	1	1997	12/1/97 17:35			
1080	GLUI	GLUCOSE (RALS)	GLUI	GLUCOSE (RALS)	GLUCOSE	(RALS)			N		N	0	0	2	2005	5/2/05 0:00			
1081	GLUMSH	GLUCOSE, BY METER (Student Health)	GLUM	GLUCOSE, BY METER	GLUCOSE	mg/dL	SH	CSBC			Student He	N	74	100	222	2006	2/2/06 0:00		
1082	GLUPOC	GLUCOSE WHOLE BLOOD (POC)	GLUPOC	GLUCOSE WHOLE BLOOD (POC)	GLUCOSE WHOLE BLOOD (POC)		N				N	0	0	22	2005	9/27/05 0:00			
1083	GLUUA	GLUCOSE, URINE	GLUUA	GLUCOSE, URINE	GLUCOSE	UR	UAS				N	0	0	491,704	2006	2/5/06 0:00			
1084	GLUUSH	GLUCOSE, URINE (Student Health)	GLUUA	GLUCOSE, URINE	GLUCOSE (URINE)	UR	UAS				N	0	0	15	2000	10/9/00 15:02			
1085	GLY	GLYCINE	GLY	GLYCINE	GLYCINE	umol/L	WEST	AAPRO			N	224	587	2,608	2006	2/1/06 0:00			
1086	GMST	GRAM STAIN	GMST	GRAM STAIN	GRAM STAIN				N		N	0	0	1	2002	2/11/02 17:00			
1087	GRASS	GRASS RAST PANEL	GRASS	GRASS RAST PANEL	GRASS RAST PANEL				N		N	0	0	12	2005	11/21/05 0:00			
1088	GRSTN	GRAM STAIN	GRSTN	GRAM STAIN	GRAM STAIN				N		N	0	0	1	2002	12/18/02 16:11			
1089	GSCPOC	GASTROCIILT (POC)	GSCPOC	GASTROCIILT (POC)	GASTROCIILT	(POC)			N		N	0	0	2	2002	9/24/02 11:00			

Example data dictionary

Definitions might change when, e.g., lab equipment changes

How to get all glucose tests? All blood glucose tests?



Controlled Terminology

- Basic features
- Unique identifier (code)
 - 12345
- Official name
 - aspirin
- Synonyms
 - ASA
- Translations
 - NDC: 5502; UMLS: C0004057



Controlled Terminology

- Advanced features
- Classes
 - Antipyretic, anti-inflammatory, analgesic
- Semantic links
 - CAUSES: gastritis
 - TREATS: arthritis
- Attributes
 - UNITS: mg



Controlled Terminologies in Clinical Care

- Diagnoses: ICD-9-CM (pre Oct 2015), ICD-10 (Oct 2015 to present), DRG
- Procedures: CPT, ICD-9 (pre Oct 2015), ICD-10 (Oct 2015 to present)
- Laboratory tests: LOINC
- Drugs: NDC, RxNorm, Multum, National Drug Formulary - Reference Terminology (NDF-RT)
- Medical Imaging: RadLex
- Clinical problems: SNOMED-CT
- Cross-references among terminologies: Unified Medical Language System (UMLS)



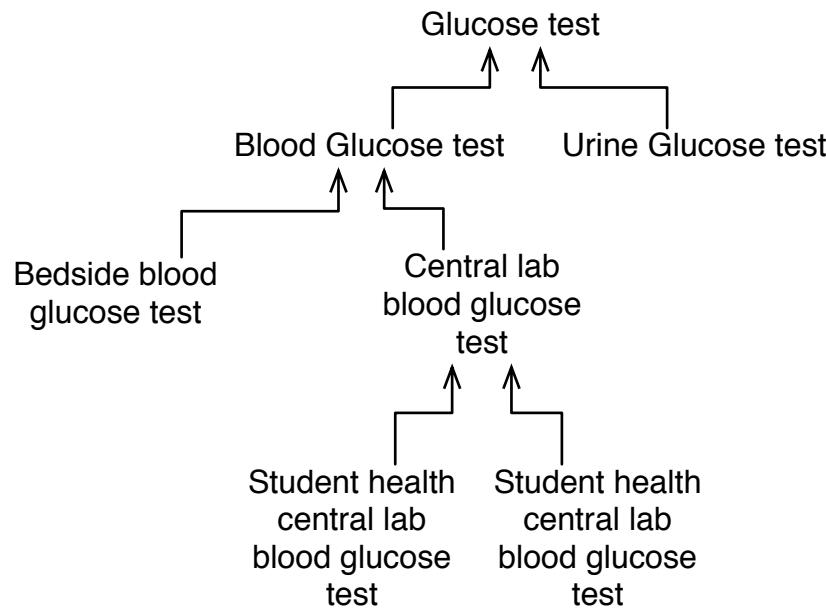
General Classes of Terms

- Procedures
 - Diagnostic
 - Therapeutic
- Medications
- Diagnoses
- Findings
- Organisms
- Anatomy



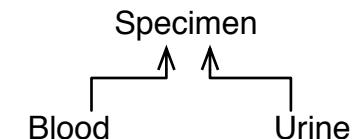
Pre- versus Post-coordinated Terms

Pre-coordinated



Post-coordinated

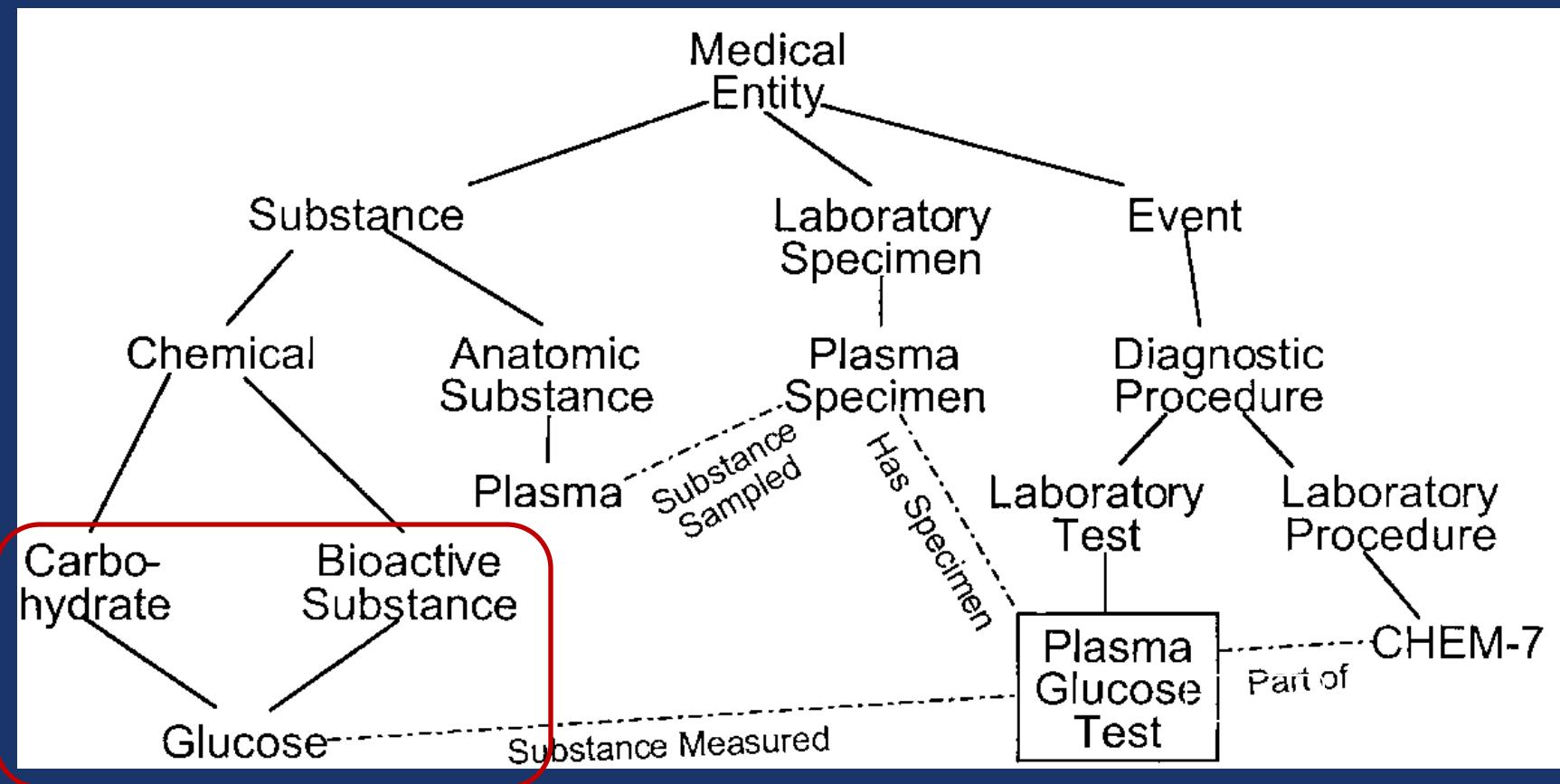
Glucose test
Patient Location
Specimen
Where analyzed



Pre-coordination makes more of a term's definition implicit in its name



Representing Medical Concepts



Concepts are defined via their semantic relationships

Concepts may belong to multiple classes



Concept-oriented Terminology

- *Concepts have no semantics, but are associated with a list of terms used to designate that concept*
- Malignant neoplasm of prostate (UMLS C0376358)
- Prostate cancer
- Cancer, prostate
- Cancers, prostate
- Prostate cancers
- Cancer of prostate
- Cancer of the prostate
- ...



Concepts Should Be

- Non-redundant
- Non-ambiguous
- Permanent
- Non-semantic



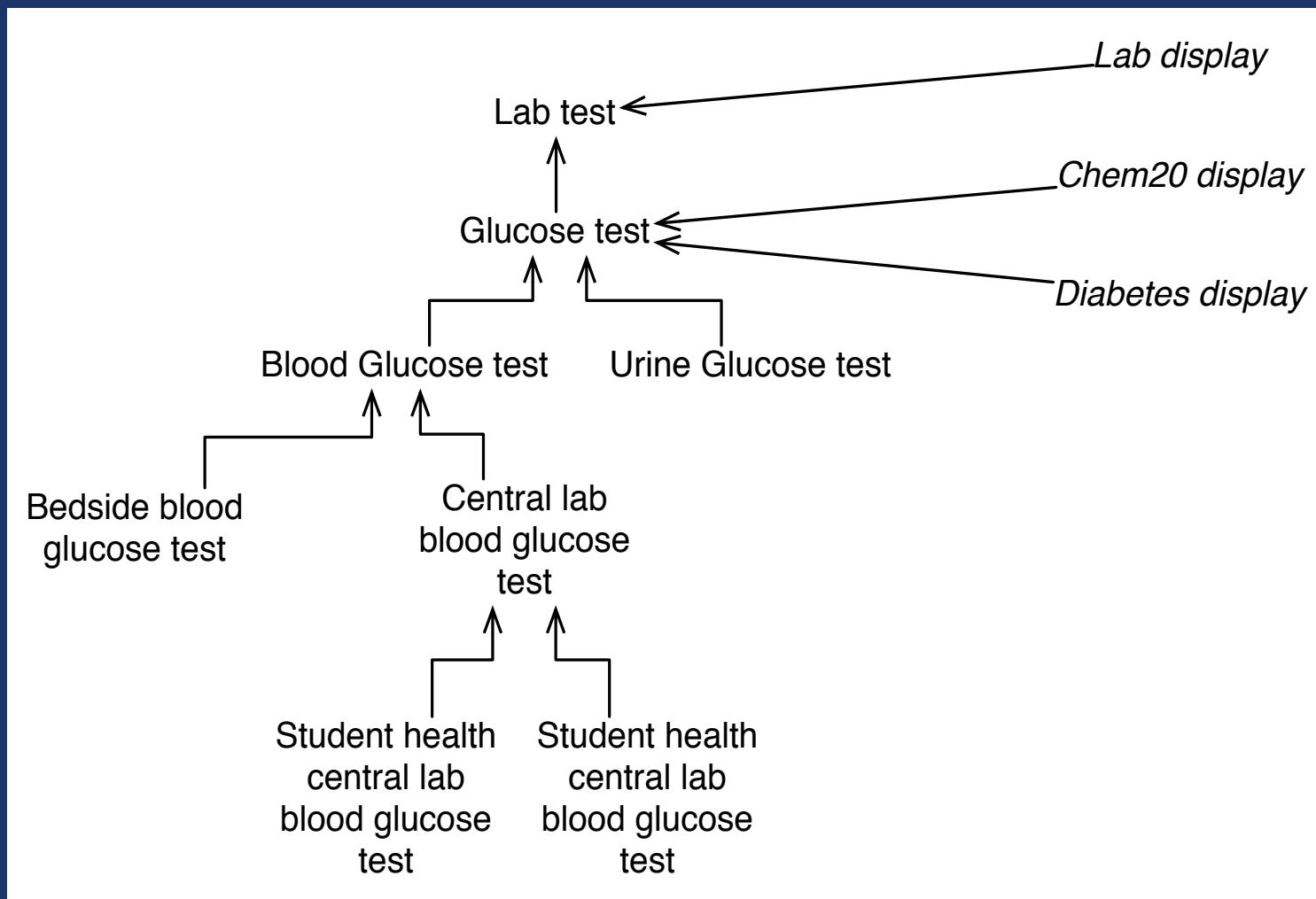
Uses of Terminologies

- Database storage and retrieval
- Data communication
- Billing
- Decision support
- Biomedical research



Clinical Data Displays

Concept-oriented Data Displays





Chem20 Display

Chem-20																				
	Na	K	Cl	HCO3	BUN	Creat	Gluc	Ca	Phos	Urate	Chol	Tot Prot	Alb	TBili	DBili	Tot Alk	PAST	ALT	LDH	CK
24Mar99 22:50	138	4.0	107	27	16	1.7	133	8.0	2.8											
25Jan99 10:00	138	4.0	105	27	11	1.2	87												165	
24Jan99 23:30																			155	
24Jan99 17:15	140	4.4	107	30	11	1.3	94				105 *		3.6	0.9	0.2	70	47	22	152	
14Oct97 13:35	142	4.6	108	28	14	1.3	85													
12Apr97 09:15	141	3.7	114	21	17	1.1	76													
10Apr97 18:20	138 *	3.8 *	111 *	26 *	28 *	1.2 *	91 *	8.0												
07Apr97 10:20	139	4.0	108	25	18	1.0	82													
06Apr97 10:10	137	3.8	107	27	13	1.0	88													
05Apr97 10:00	141	3.7	107	26	11	1.0	87													
	Na	K	Cl	HCO3	BUN	Creat	Gluc	Ca	Phos	Urate	Chol	Tot Prot	Alb	TBili	DBili	Tot Alk	PAST	ALT	LDH	CK
04Apr97 13:08	144	4.1	112	26			107													
24Mar97 15:12	143	4.0	109	32	12	1.2	74					7.0	3.5	0.4	0.1	74	26	25		
14Dec96 21:13	SPE	SPE	SPE	SPE	SPE	SPE	SPE													
13Dec96 12:50	138	3.5	110	28	7	1.0	80													
12Dec96 12:10	140 *	3.9 *	108 *	25 *	6 *	1.0 *	82 *	8.3	1.6											
11Dec96 04:40	144 *	3.7 *	114 *	24 *	11 *	0.9 *	99 *	8.1	2.1				3.2		41	22	9	94 *		
10Dec96 21:20	145	3.8	113	24	13	0.9	84													

Cimino JJ. From data to knowledge through concept-oriented terminologies: Experience with the medical entities dictionary. J Am Med Inform Assoc 2000, May;7(3):288-97.



Concept-oriented Display

A query for pulmonary heart disease yields

- Test results that measure relevant substances (e.g., O₂, CO₂)
- Reports on exams of relevant body parts (e.g., cardiograms, chest x-rays)
- Medications that treat this disease

[View by Department](#) [View by Time](#) [View by Topic](#)

For Tests/Reports Related to Pulmonary Heart Disease

Search among: [Lab Reports](#) | [Radiology Reports](#) | [Adission/Discharge Diagnosis](#) | [Drug Orders](#)
Browse : [Cardiology Reports](#)

Pharmacy	
98/07/17 10:00	UD Capoten 12.5 mg Tab D <input checked="" type="checkbox"/>
98/07/17 10:00	UD Isosorbide Dinitrate 10 mg D <input checked="" type="checkbox"/>
98/07/04 16:00	UD Capoten 25 mg Tab D <input checked="" type="checkbox"/>
98/07/04 10:00	UD Isosorbide Dinitrate 10 mg D <input checked="" type="checkbox"/>
98/06/23 10:00	UD Isosorbide Dinitra 5mg Oral D <input checked="" type="checkbox"/>
98/06/23 10:00	Heparin 5000 U/ml 10 ml D <input checked="" type="checkbox"/>
.....	UD Isosorbide Dinitra 5mg D <input checked="" type="checkbox"/>

Oral Medication Order	
Ordering Physician	VALERI, ANTHONY M
CPMC Drug: UD Capoten 12.5 mg Tab	B
Pharmacy Order Route	PO
Pharmacy Order Drug Dose	12.500000 MG
Pharmacy Order Final Concentration	12.500000 MG
Pharmacy Order Drug Strength	12.500000 MG
Pharmacy Order Effective Time	98/07/21 14:44
Patient Name	UNKNOWN NAME



Data Translation

- Local data is stored in milliliters
- A decision support rule uses liters
- Program logic can translate the local data into the appropriate form for uniform display and analysis of data from heterogeneous sources



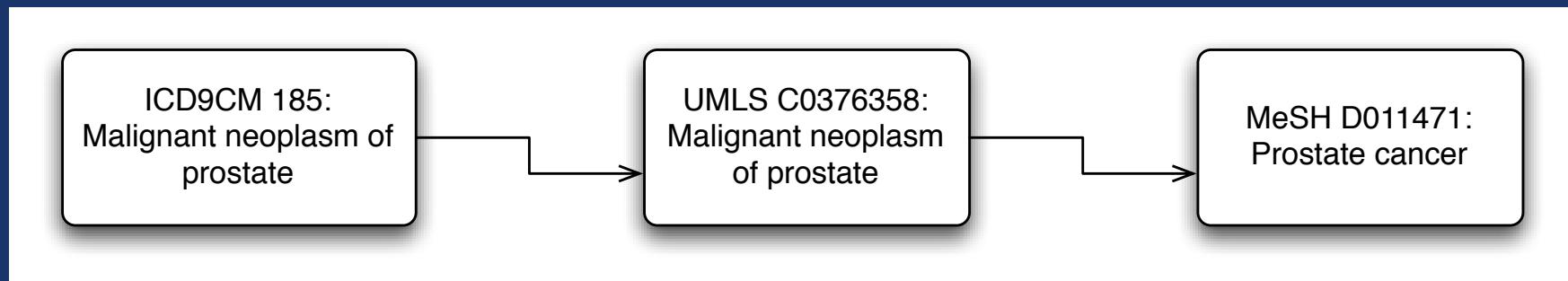
Decision Support with Terminologies

- User interface creation
 - Store data view definitions as concepts in a terminology with links to the data they present
 - Suggest other data elements that are missing from the view
 - Find data elements that are not presented in any view



Decision Support with Terminologies

InfoButtons



Translate diagnosis and procedure codes from the patient into the vocabularies of various information sources (e.g., MeSH)



Decision Support with Terminologies

Rule-based/guideline-based Systems

- High-level concepts (e.g., Negative blood culture) specified in rules can be integrated with the low-level concepts/free text found in EMRs (e.g., “No growth”, “No growth in 24 hours”)
- Rules will continue to function when changes occur to how tests are performed
- Free-text processing is facilitated by searching for the terms associated with a concept



Example: TB

- A rule monitors for positive culture results
- Sends message if result not equal to “No growth”
- One day, dozens of alerts about positive results but no organism was reported
- What happened?



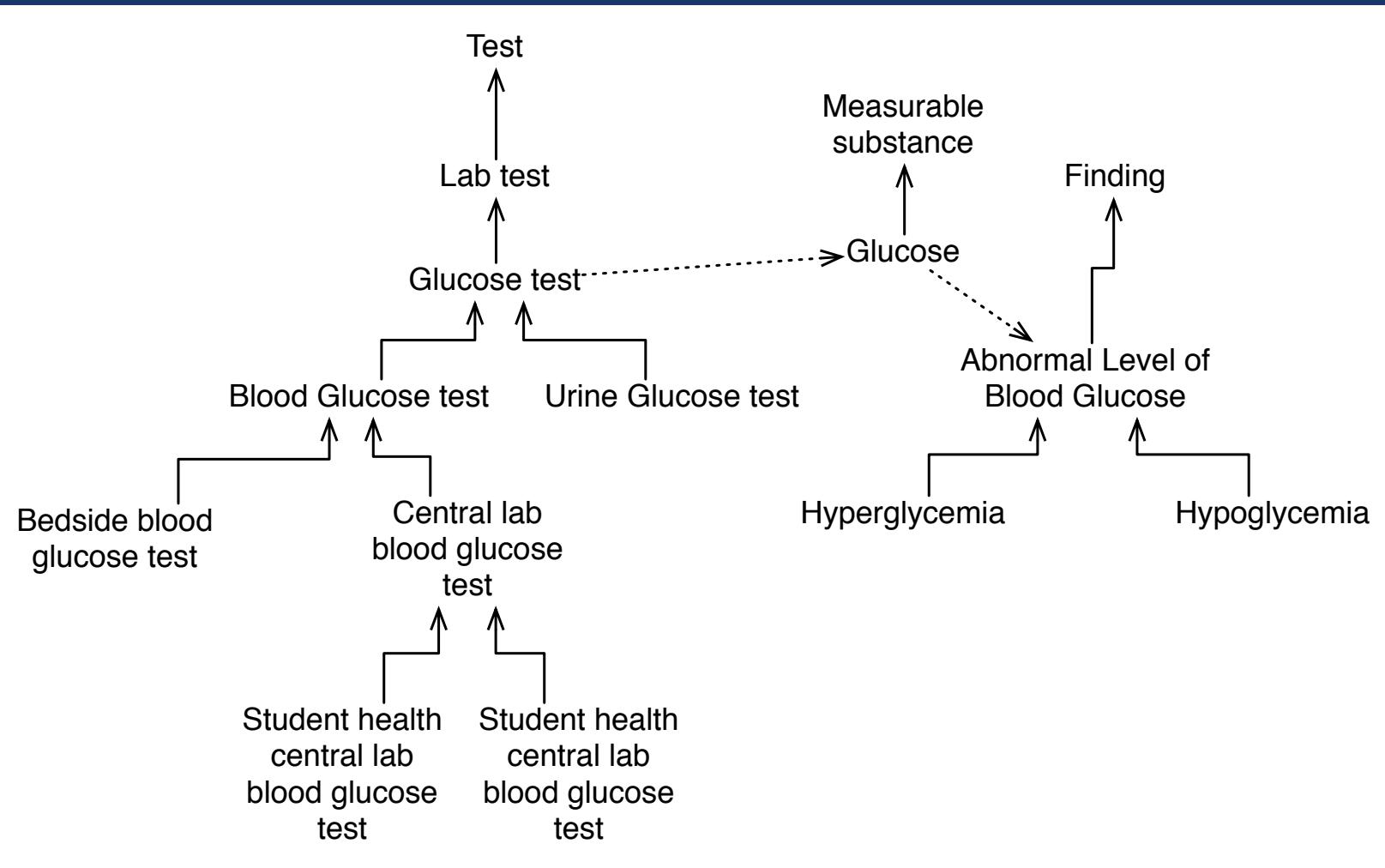
Example: TB

- Alert looked for specific text (“No growth”)
- Lab started reporting “No growth to date”)
- “No growth” <> “No growth to date”
- We could either update every software module that relies on the phrase “No growth” being there, or...
- We could create a NoGrowth concept class that stores all corresponding terms, and have the rule (and all other users of “No growth” refer to the class)



Decision Support with Terminologies

Translating Data into Findings





Decision Support with Terminologies

Diagnostic Systems

Below is the list of laboratory tests and findings ALREADY included in DXplain's search request:

- | | | |
|---|---|---|
| <input checked="" type="checkbox"/> ELDERLY (>65 YRS) | <input checked="" type="checkbox"/> MALE | <input checked="" type="checkbox"/> Hyperglycemia |
| <input checked="" type="checkbox"/> Hyponatremia | <input checked="" type="checkbox"/> Hypokalemia | <input checked="" type="checkbox"/> Hypoalbuminemia |
| <input checked="" type="checkbox"/> Creatinine, Elevated | <input checked="" type="checkbox"/> Hypocalcemia | <input checked="" type="checkbox"/> Serum Conjugated Bilirubin Elevated |
| <input checked="" type="checkbox"/> Serum Total Bilirubin Elevated | <input checked="" type="checkbox"/> Sgot (Alt), Elevated | <input checked="" type="checkbox"/> Alkaline Phosphatase, Elevated |
| <input checked="" type="checkbox"/> Sgot (Ast), Elevated | <input checked="" type="checkbox"/> no Bicarbonate, Increased | <input checked="" type="checkbox"/> no Hyperchloremia |
| <input checked="" type="checkbox"/> no Hypochloremia | <input checked="" type="checkbox"/> no Blood Urea Nitrogen Elevated | <input checked="" type="checkbox"/> no Bicarbonate, Decreased |
| <input checked="" type="checkbox"/> no Blood Urea Nitrogen Decreased | <input checked="" type="checkbox"/> no Hypouricemia | <input checked="" type="checkbox"/> no Serum Phosphate Decreased |
| <input checked="" type="checkbox"/> no Serum Phosphate Elevated | <input checked="" type="checkbox"/> no Serum Creatine Phosphokinase | <input checked="" type="checkbox"/> no Hyperuricemia |
| <input checked="" type="checkbox"/> no Serum Lactic Acid Dehydrogenase Elevated | Elevated | |

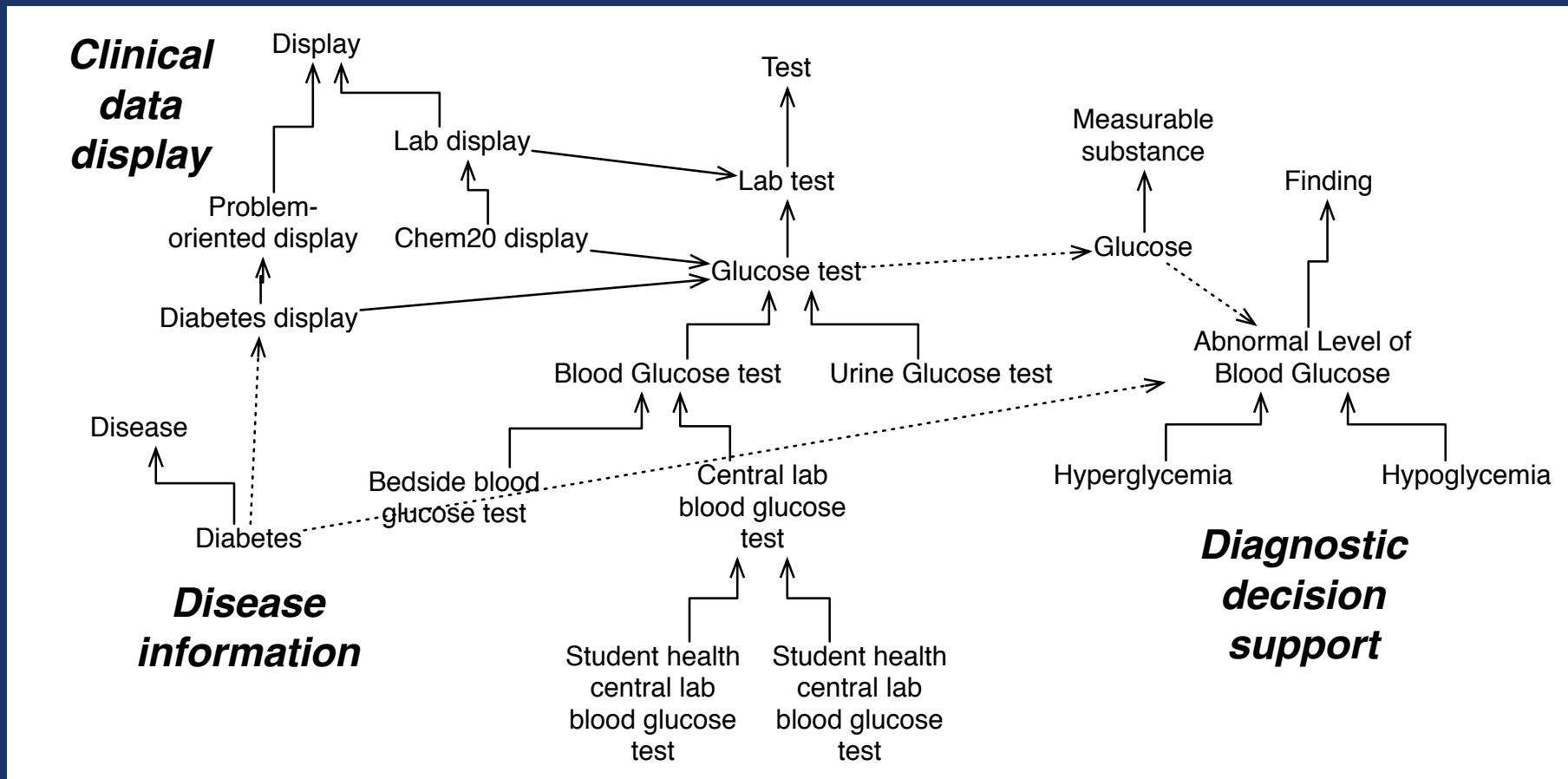
To see the text associated with each diagnosis hit the button on the left, when you are done hit [back]

DXplain's Diagnoses		Interp.
	Disease Information	Explain Disease
1	ALCOHOLISM	+
2	DIABETES MELLITUS, NON-INSULIN DEPENDENT	+
3	MAGNESIUM DEFICIENCY SYNDROME	+
4	COLITIS, ULCERATIVE	+
5	NON-KETOTIC HYPEROSMOLAR COMA	+
6	RENAL CELL CARCINOMA	+
7	NEPHROTIC SYNDROME	
8	CHOLECYSTITIS, ACUTE	
9	ENTERITIS, REGIONAL (CROHN'S DISEASE)	
10	HEART FAILURE, CONGESTIVE	

Cimino JJ. From data to knowledge through concept-oriented terminologies: Experience with the medical entities dictionary. J Am Med Inform Assoc 2000, May;7(3):288-97.

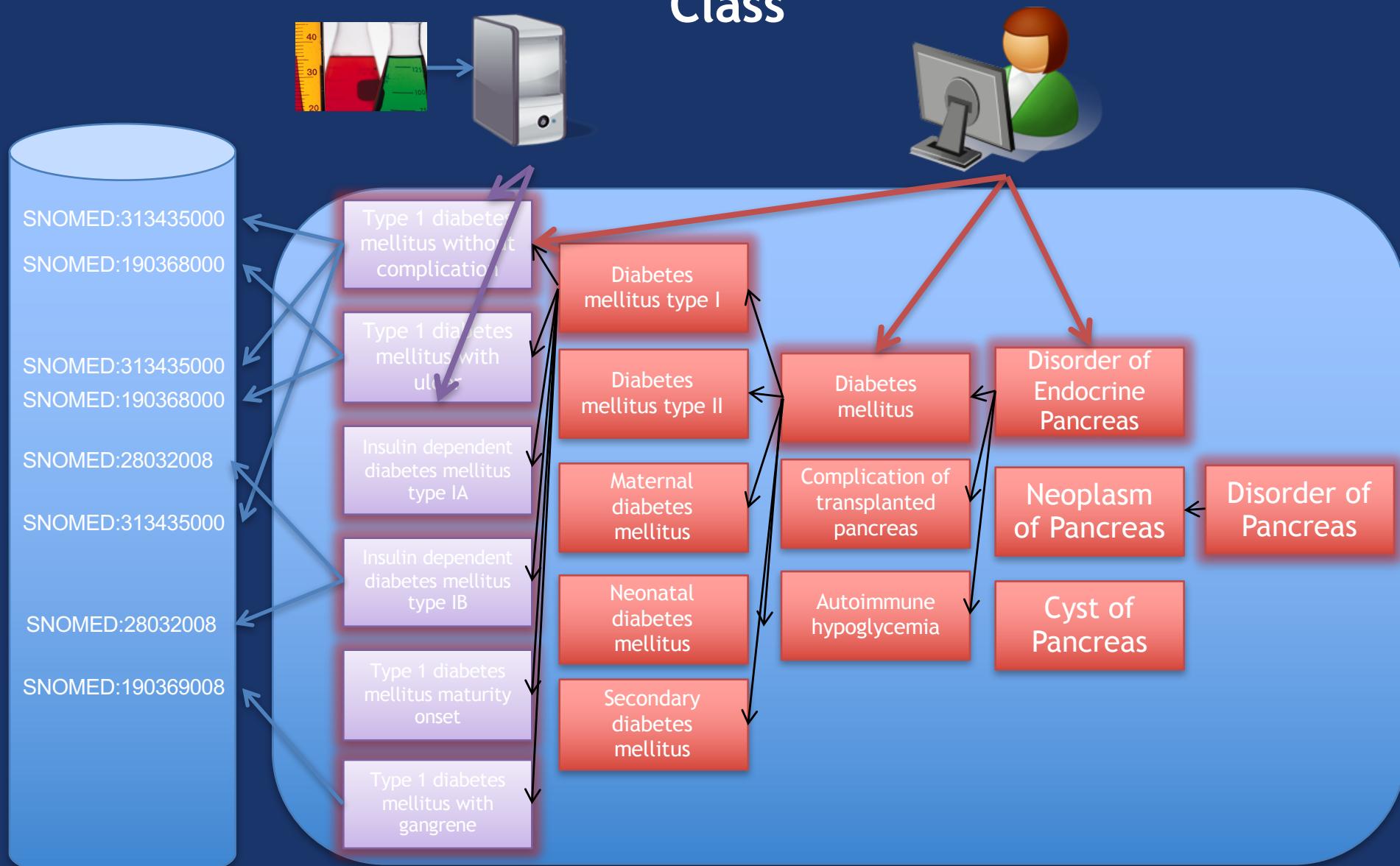


The Promise: A Controlled Terminology Can Support Many Applications through Sharing and Reuse





Coded Concepts/Terms Support Retrieval of Data by Class





Coded Concepts/Terms Support Aggregation of Semantically-related Data for Analysis



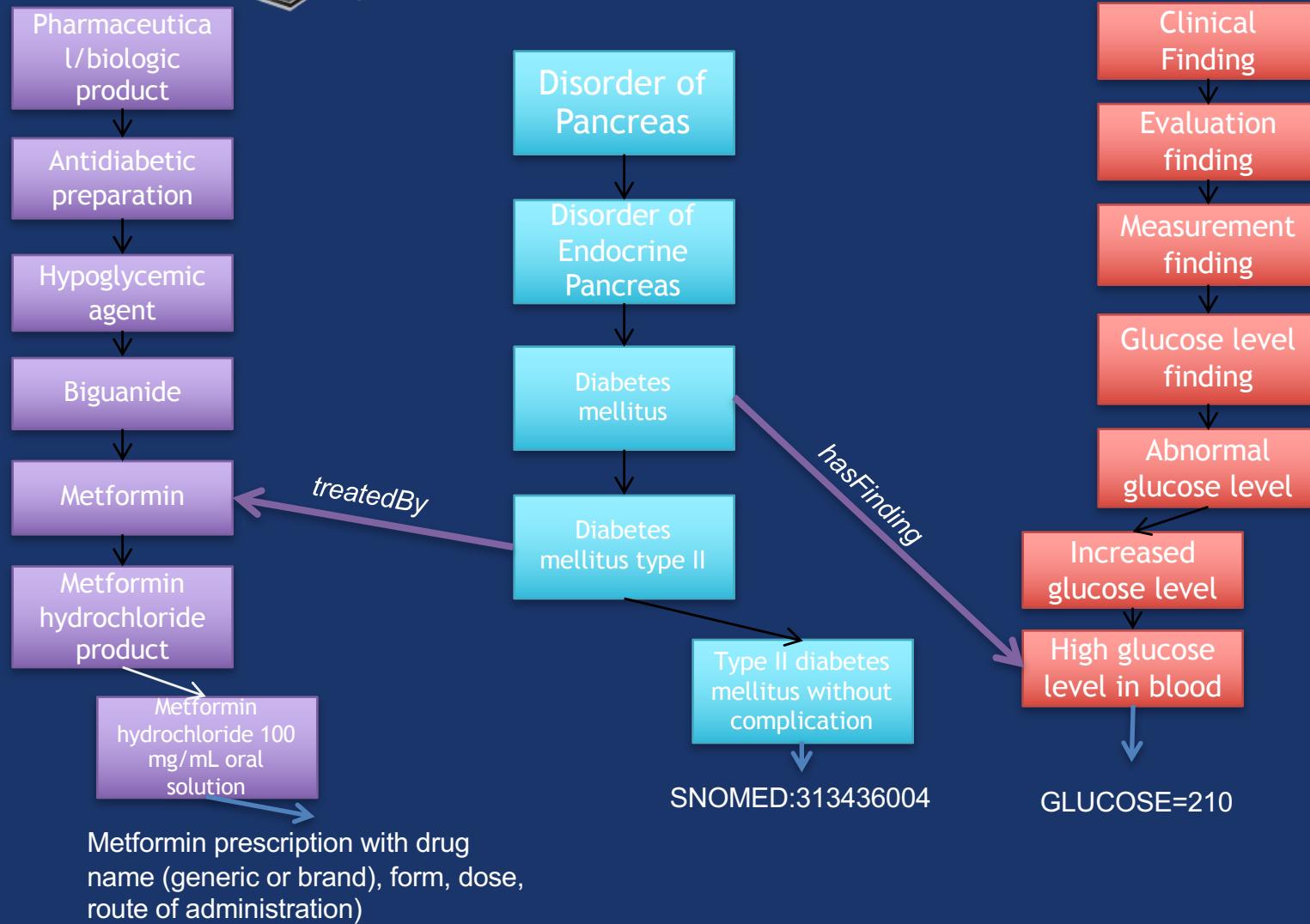
“Get me all patients with (list of type II diabetes diagnosis codes), x number of (codes for a blood glucose test) > threshold or (list of codes for medications that treat type II diabetes)”



Coded Concepts/Terms Support Aggregation of Semantically-related Data for Analysis



"Get me all patients with type II diabetes as evidenced by diagnoses, lab test results and medication history"





Past Reality: Shareable Controlled Terminologies are not Commonly in Use

- Rare success stories
 - Drug formularies
 - Drug-drug interaction data
- Successes are simple
 - Simple tables that can be loaded into a relational database (e.g., Micromedex)
 - Free-text information resources that do not contain executable knowledge (e.g., PubMed, clinicaltrials.gov)



Why?

- Could controlled terminologies' potential benefits of sharing and reuse be substantial drivers for their adoption?
 - Not so far
 - Terminologies are extremely time consuming and expensive to construct, especially if the goal is reusability
 - If the terminology needs to be extended for every new application, what are the benefits really?
 - The need for communication between vendors has been a driver for terminologies/standardization in other industries, but few business cases exist for this kind of communication in medicine except for reimbursement



Present/Future Reality: Meaningful Use/MACRA

- Stage 1: Adopt EHRs and store data in structured form (starting 2011)
- Stage 2: Begin advanced care processes with decision support and care coordination (starting 2014)
- Stage 3: Show improved outcomes and reduced costs through defined quality measures (starting 2017)
- Reimbursement bonuses from Medicare and Medicaid for physicians who attest successfully
- Reimbursement penalties for those who do not starting in 2015



National Quality Forum

Log In My Dashboard Find Measures Calendar About Us Keyword Search

 NATIONAL QUALITY FORUM

Measures (Result List) Portfolios Compare Add to Compare Add to Portfolio Export ? Give Feedback

0421 Preventive Care and Screening: Body Mass Index (BMI) Screening and Follow-Up
STEWARD: Centers for Medicare & Medicaid

Measure Description:
Percentage of patients aged 18 years and older with a documented BMI during the current encounter or during the previous six months AND when the BMI is outside of normal parameters, a follow-up plan is documented during the encounter or during the previous six months of the encounter.
Normal Parameters: Age 65 years and older BMI > or = 23 and < 30
Age 18 – 64 years BMI > or = 18.5 and < 25

Numerator Statement:
Patients with a documented BMI during the encounter or during the previous six months, AND when the BMI is outside of normal parameters, follow-up is documented during the encounter or during the previous six months of the encounter with the BMI outside of normal parameters

Denominator Statement:
All patients aged 18 years and older

Exclusions:
A patient is identified as a Denominator Exclusions (B) and excluded from the Total Denominator Population (TDP) in the Performance Denominator (PD) calculation if one or more of the following reason (s) exist:

- Patient is receiving palliative care
- Patient is pregnant
- Patient refuses BMI measurement (refuses height and/or weight)
- Any other reason documented in the medical record by the provider why BMI calculation or follow-up plan was not appropriate
- Patient is in an urgent or emergent medical situation where time is of the essence, and to delay treatment would jeopardize the patient's health status

Risk Adjustment:
No

Classification:
National Quality Strategy Priorities:
Health and Well-Being

Measure Steward Contact Information:
For additional measure specification information, please contact the Measure

Take Action

Status
Endorsement Type:
Endorsed
Last Updated Date:
Mar 12, 2014
eMeasure Available:
Yes
Measure History:
Full History

Found in Portfolio(s)
[2012 MAP Cardiovascular Family of Measures](#)
[2012 MAP Diabetes Family of Measures](#)
[ACO - PCMH Measures](#)
[ACO Measures](#)
[Chris ACO measures](#)
[Disparities-sensitive Measures](#)
[Health and Well-Being](#)
[Healthy York County Coalition Publicly Reported Measures \(South Central PA AF4Q\)](#)
[Maine Health Management Coalition](#)



Agenda

- Data management landscape
- Data management systems
- The case for structured data
- Terminologies/vocabularies/etc
- **Intersection with machine learning**

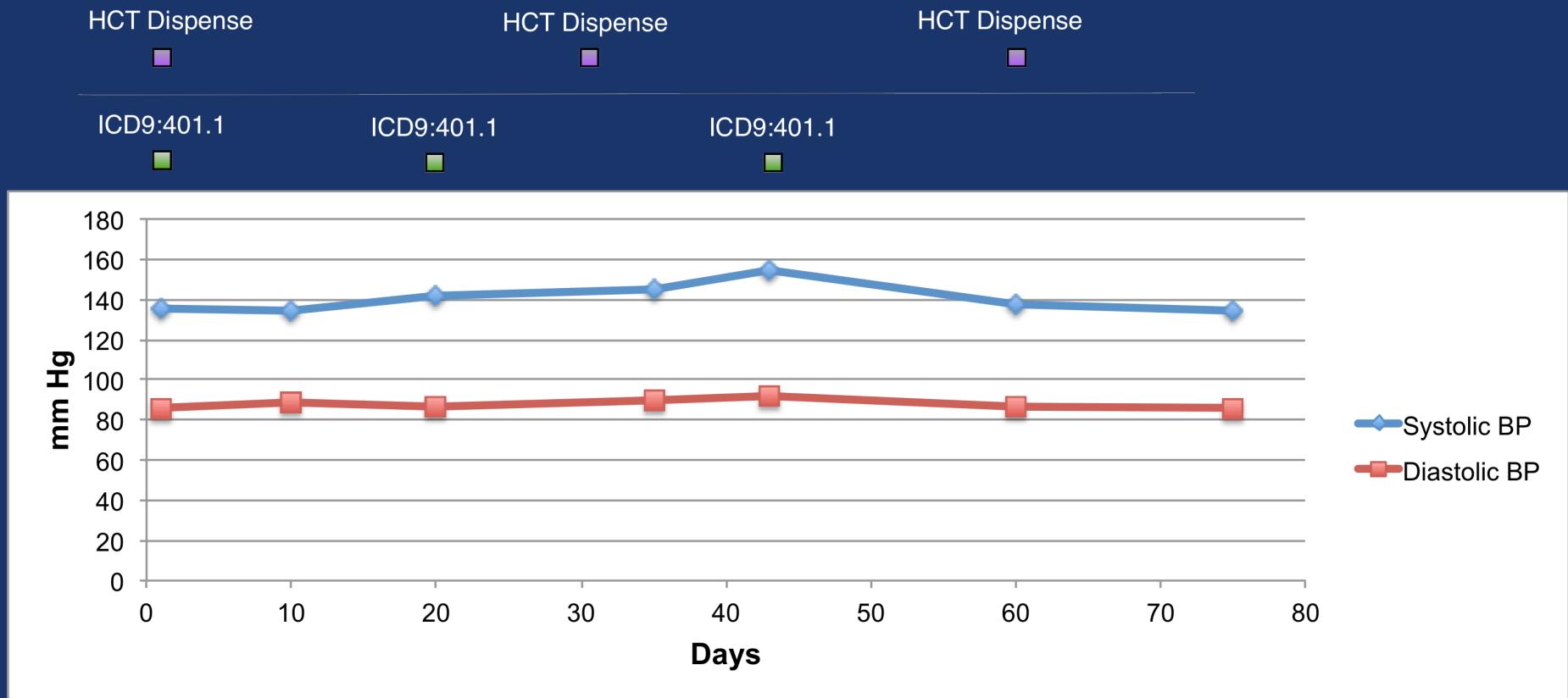


Challenges in Using EHR Data for Prediction

- Problem: “Raw” variables in clinical data warehouses are difficult to use for predictive modeling
 - Too many diagnosis codes, procedure codes
 - Continuous variables (e.g., labs) require interpretation
 - Temporal relationships between variables are implicit
- Solution: Transform the data into a much smaller set of variables using heuristic knowledge
 - Categorize diagnosis and procedure codes using code hierarchies
 - Classify continuous variables using standard interpretations (e.g., high, normal, low)
 - Identify temporal patterns (e.g., frequency, duration, sequence)
 - Apply standard data mining techniques



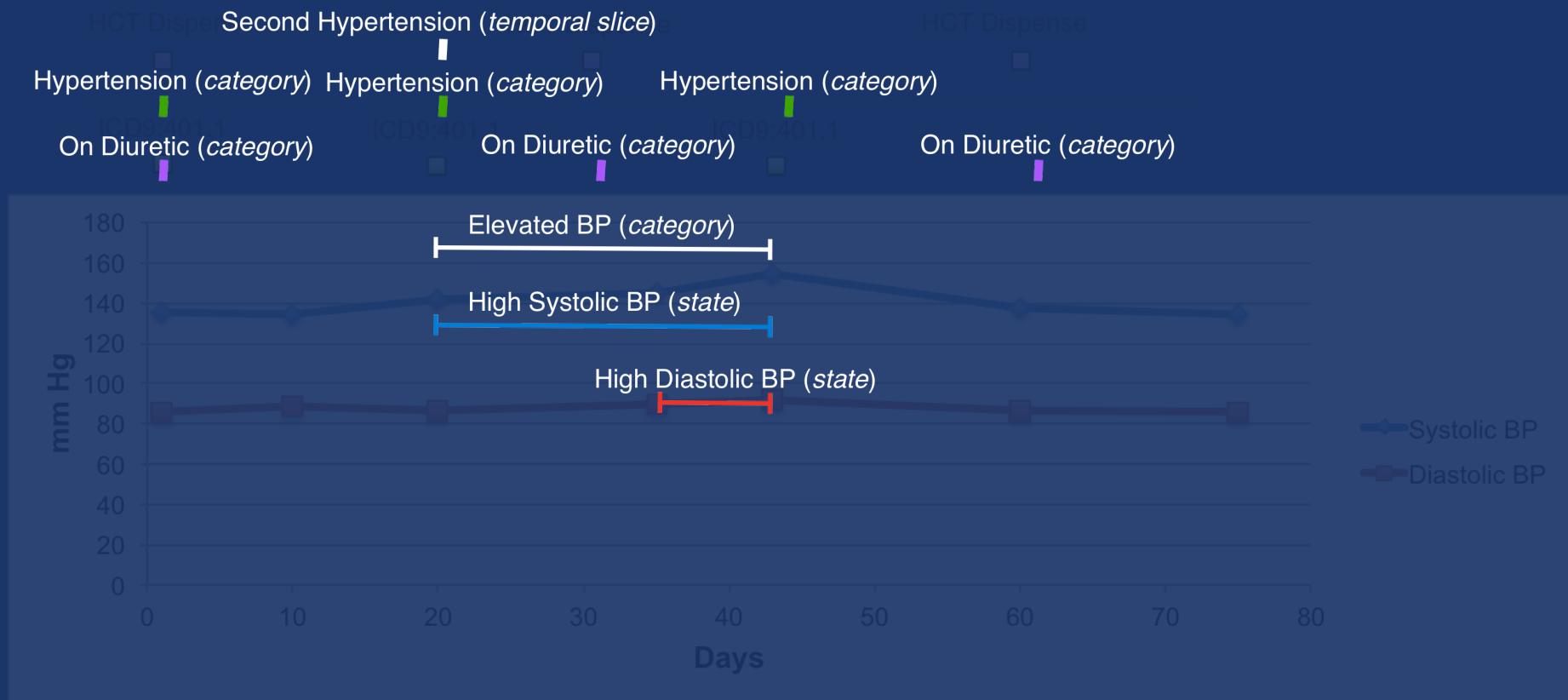
From Snapshots to Patient Histories



Post AR & Harrison JH. J Am Med Inform Assoc 2007;14(5):674-83



Abstraction of Data into Intervals





Eureka! Clinical Analytics

<http://eurekaclinical.org>

Specify patient features of interest

Phenotype Editor

+ Create New Element

Action	Name	Description	Type	Created Date	Last Modified
	High Blood Pressure	High blood pressure. Includes ranges for all patients and patients with CKD or diabetes.	Value Threshold	02/06/2013	07/31/2013
	2 consecutive high blood pressure values within 180 days	2 consecutive high blood pressure values within 180 days.	Frequency	02/06/2013	02/15/2013
	Digestive procedure with prior chemo	Digestive procedure with prior chemo	Sequence	02/14/2013	02/14/2013
	Readmission	A readmission.	Frequency	02/15/2013	02/15/2013

Compute them in millions of patient records

Elev. BP in Hypertensive on Diuretic (temporal pattern)

Second Hypertension (temporal slice)

Hypertension (category) Hypertension (category) Hypertension (category)

On Diuretic (category) On Diuretic (category) On Diuretic (category)

Elevated BP (category)

High Systolic BP (state)

High Diastolic BP (state)

time

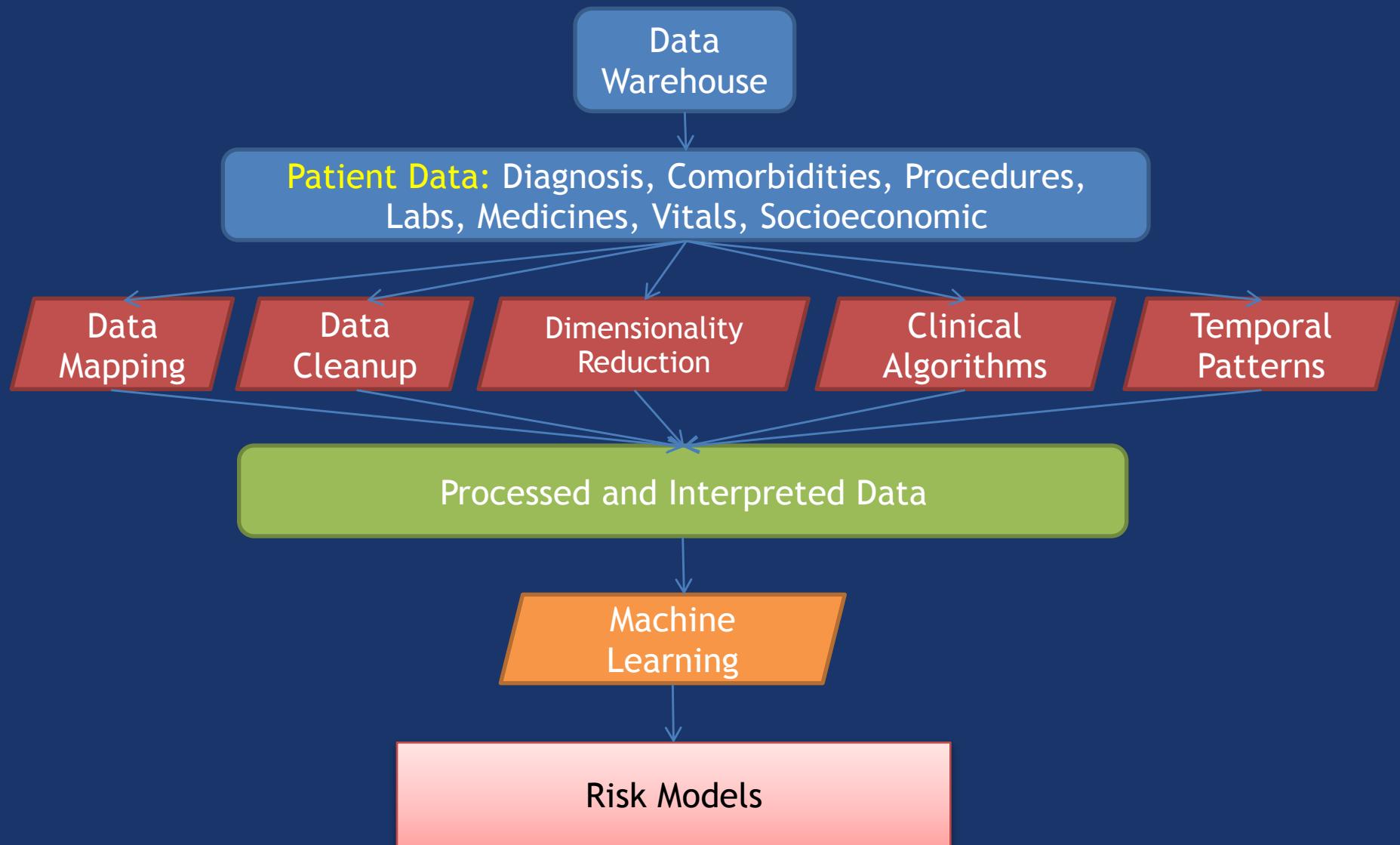
Spreadsheets

Csv

Clinical databases



HPA Workflow





Healthcare Predictive Analytics

Opportunities

- High-cost patients
- Readmissions
- Triage
- Decompensation
- Adverse events
- Treatment optimization for diseases affecting multiple organ systems



READ
SOMETHING
THAT MEANS
SOMETHING.
\$1 A WEEK >

SUBSCRIBE
RENEW
GIVE A GIFT
NON-U.S.
ORDERS

[Sign in](#) | [Link your subscription](#) | [TNY Store](#)



THE NEW YORKER

NEWS

CULTURE

BOOKS & FICTION

SCIENCE & TECH

BUSINESS

HUMOR

MAGAZINE

ARCHIVE

SUBSCRIBE



MEDICAL REPORT | JANUARY 24, 2011 ISSUE

THE HOT SPOTTERS

Can we lower medical costs by giving the neediest patients better care?

BY ATUL GAWANDE



TABLE OF
CONTENTS

If Camden, New Jersey, becomes the first American community to lower its medical costs, it will have a murder to thank. At nine-fifty on a February night in 2001, a twenty-two-year-old black man was shot while driving his Ford Taurus station wagon through a neighborhood on the edge of the Rutgers University campus. The victim lay motionless in the street beside the open door on the driver's side, as if the car had ejected him. A neighborhood couple, a physical therapist and a volunteer firefighter, approached to see if they could help, but police waved them back.

"He's not going to make it," an officer reportedly told the physical therapist.



*In Camden, New Jersey, one per cent of patients account for a third of the city's medical costs.
Photograph by Phillip Toledano.*

SIGN UP FOR NEWSLETTERS

E-mail address

SIGN UP

ADVERTISEMENT

YOU KNOW
EXACTLY
WHAT YOUR
BUSINESS NEEDS



Hospital Readmissions within 30 Days

5 years of data from local data warehouse - 230,000 hospital encounters

5 years of data from the UHC Clinical Database -- ~20 million hospital encounters

- Classify inpatient encounters into high, medium, low risk groups of 30-day readmission based on patients' characteristics
- Clinically meaningful derived variables
- Predictive modeling
 - Random forests (ensemble of decision trees)
 - Ranking of the predictions into high to low risk



Logistic Regression Analysis of Selected Derived Variables

Disease category	Derived variable	Level	Odds Ratio	p value
Cancer	Metastasis	Yes vs. No	1.331	0.001
	Multiple MIs	Yes vs. No	2.188	< 0.0001
Transplant	Pressure Ulcer	Yes vs. No	3.324	0.004
Myocardial Infarction	Heart Failure from BNP	Suggest_heart_failure_is_present vs. No_heart_failure	1.096	0.0289
		Indicate_mild_heart_failure vs. No_heart_failure	1.205	
		Indicate_moderate_heart_failure vs. No_heart_failure	1.295	
		Indicate_severe_heart_failure vs. No_heart_failure	1.422	
Sickle Cell Anemia	County Average House Value	<= \$87,400 vs > \$140,000	1.594	< 0.0001
		\$87,400 - \$101,800 vs > \$140,000	2.153	
		\$101,800 - \$140,000 vs > \$140,000	1.867	
Chronic Kidney Disease	End-stage Renal Disease	End-stage Renal Disease vs. Chronic Kidney Disease	1.395	< 0.0001



Random Forests

- Random forests: an ensemble of decision tree predictors
- Each tree is created using a random subset of the variables in the dataset
- A large number of trees are generated
- All of them vote to classify a test example



On-admission Prediction of Subsequent Readmission within 30 Days using Random Forest Algorithm

Diagnosis Category	Ratio of Readmission Rate in Predictive Modeling Group versus Baseline Readmission Rate	
	Predictive Modeling High Risk Group	Predictive Modeling Low Risk Group
Cancer	1.9783	0.3018
Chronic kidney disease	1.8473	0.3744
Chronic obstructive pulmonary disease	2.0979	0.4911
Diabetes	2.0186	0.3656
Heart failure	2.0123	0.3775
Myocardial infarction	2.1953	0.3468
Pulmonary hypertension	1.8315	0.4410
Sickle cell anemia	1.5548	0.2376
Stroke	2.4180	0.3589
Transplant	1.7573	0.4610



Post-discharge Prediction of Subsequent Readmission within 30 Days using Random Forest Algorithm

Diagnosis Category	Ratio of Readmission Rate in Predictive Modeling Group versus Baseline Readmission Rate	
	Predictive Modeling High Risk Group	Predictive Modeling Low Risk Group
Cancer	2.6610	0.2392
Chronic kidney disease	2.2941	0.2448
Chronic obstructive pulmonary disease	2.5874	0.2934
Diabetes	2.6659	0.2337
Heart failure	2.4869	0.2298
Myocardial infarction	2.9349	0.1147
Pulmonary hypertension	2.2256	0.1459
Sickle cell anemia	2.0678	0.2496
Stroke	3.9265	0.1667
Transplant	2.1347	0.4014



Recap

- Data management landscape
- Data management systems
- The case for structured data
- Terminologies/vocabularies/etc
- Intersection with machine learning



EMORY
UNIVERSITY

Department of
Biomedical Informatics



QUESTIONS?