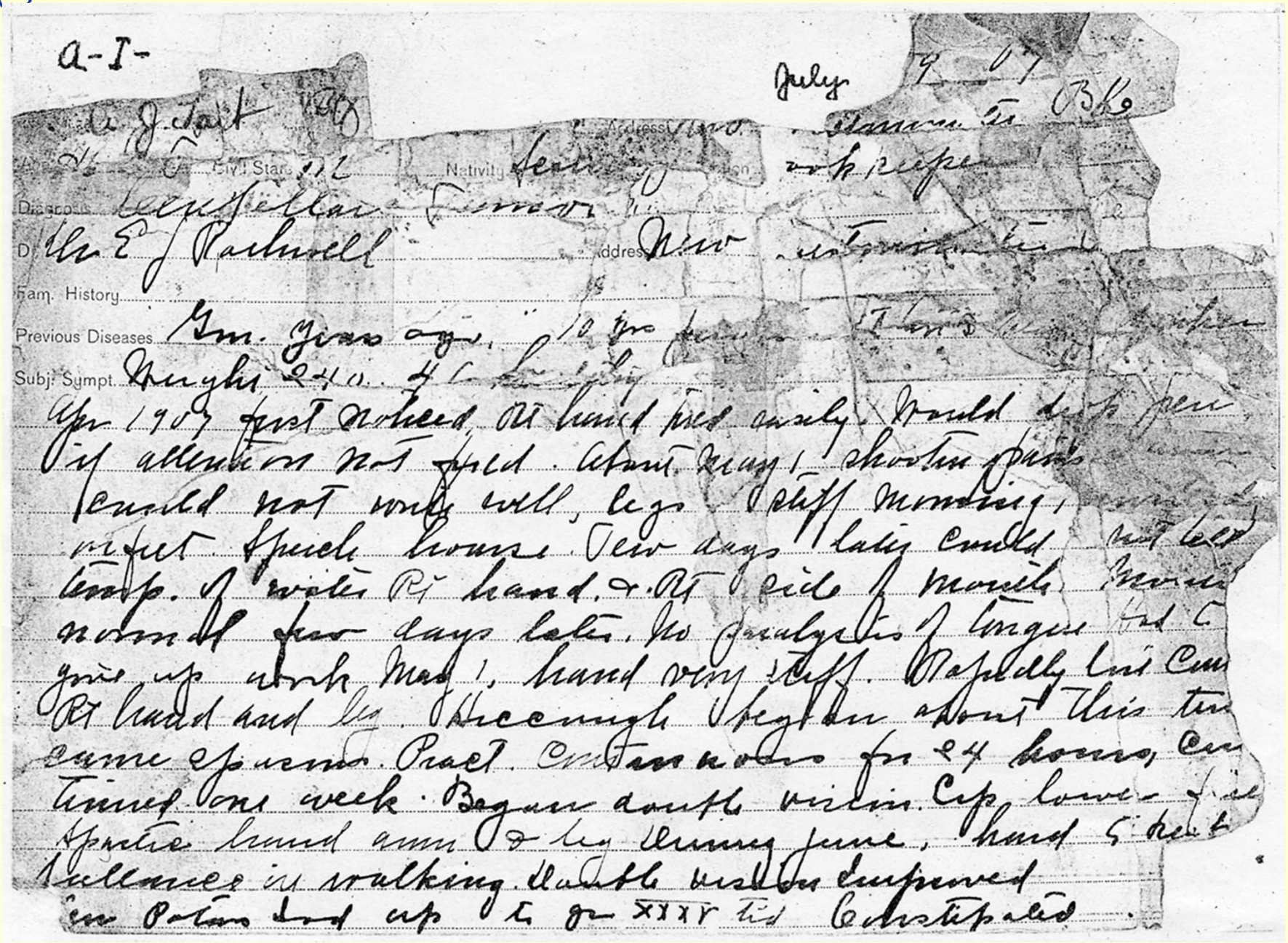
Strategic Health IT Advanced Research Projects (SHARP): Secondary Data Use and Comments on Mayo Preservation

Christopher G. Chute, MD DrPH
Mayo Clinic

NLM/NIST: Long-term Preservation of EMR

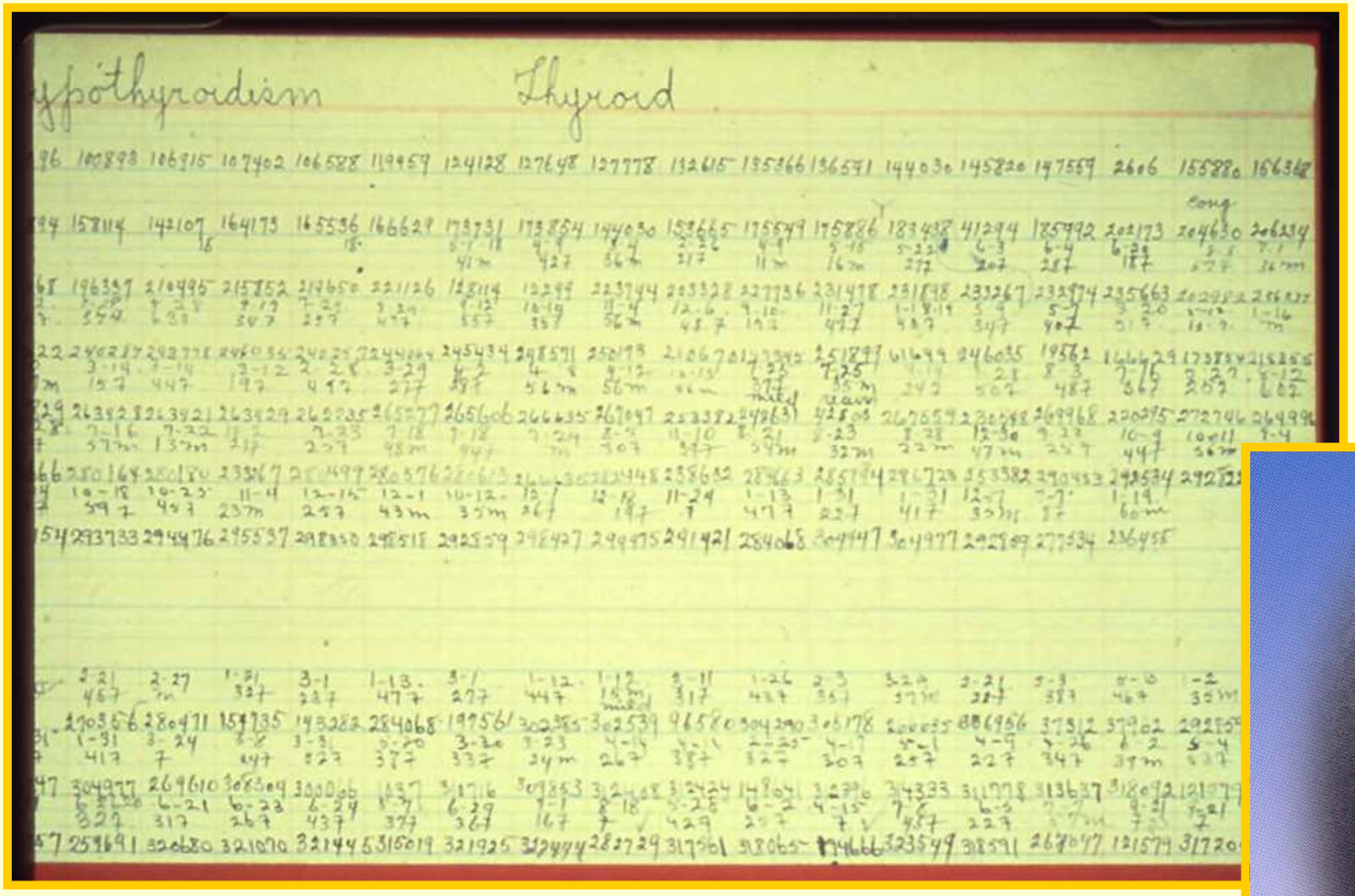
6 April 2011

Original Mayo Medical Record Ledgers 1885 - 1907



1907: First Mayo Clinic Medical Record

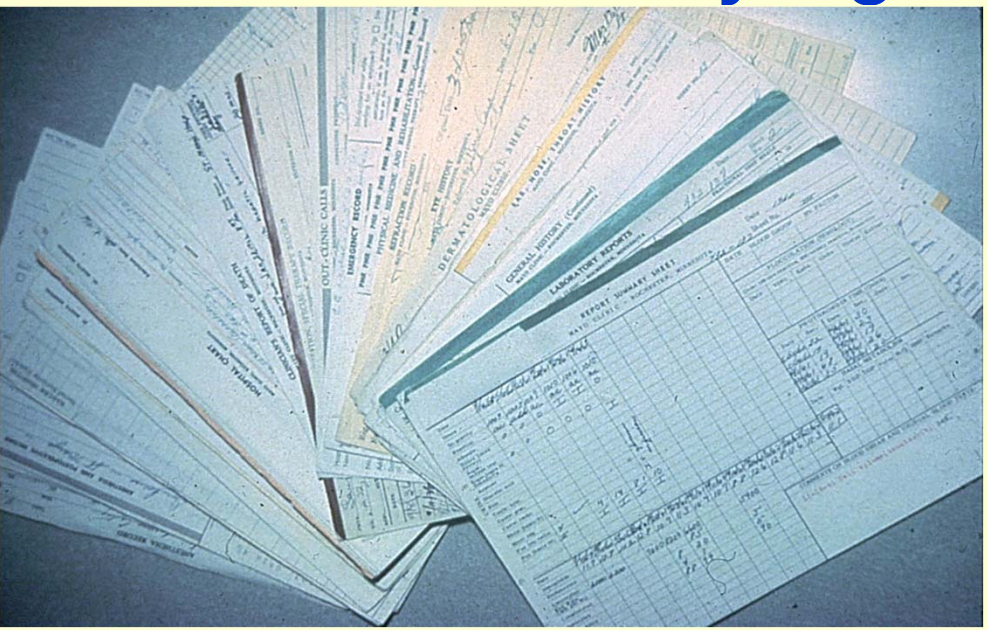
1910: Mabel Root



- **Card catalog**
- **5 x 8 inch index cards**
- **Summary for research**

Plummer-Root Original Six Wooden Medical Record Cabinets

Mayo: A Century-Long Tradition of Studying Patient Outcomes

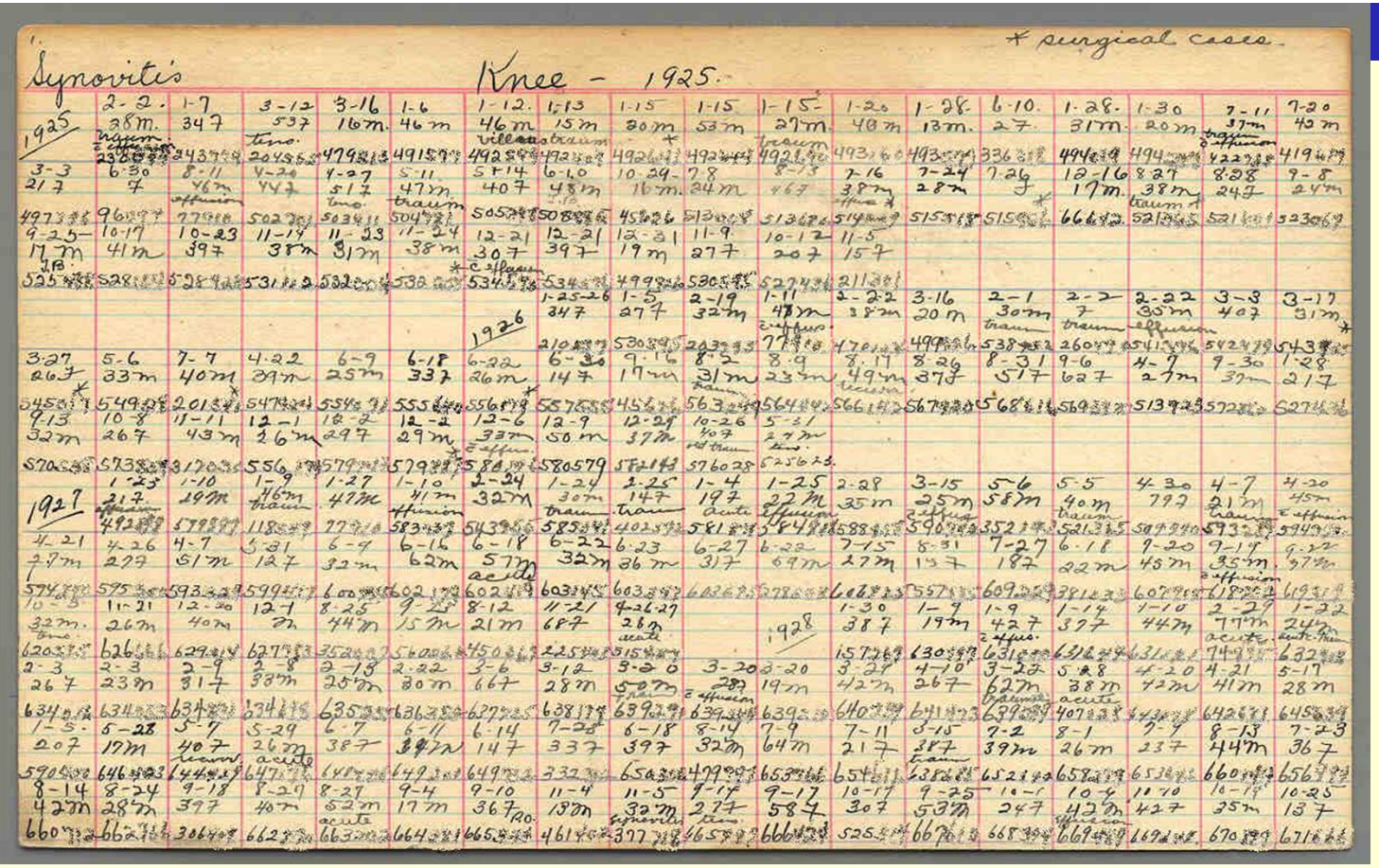


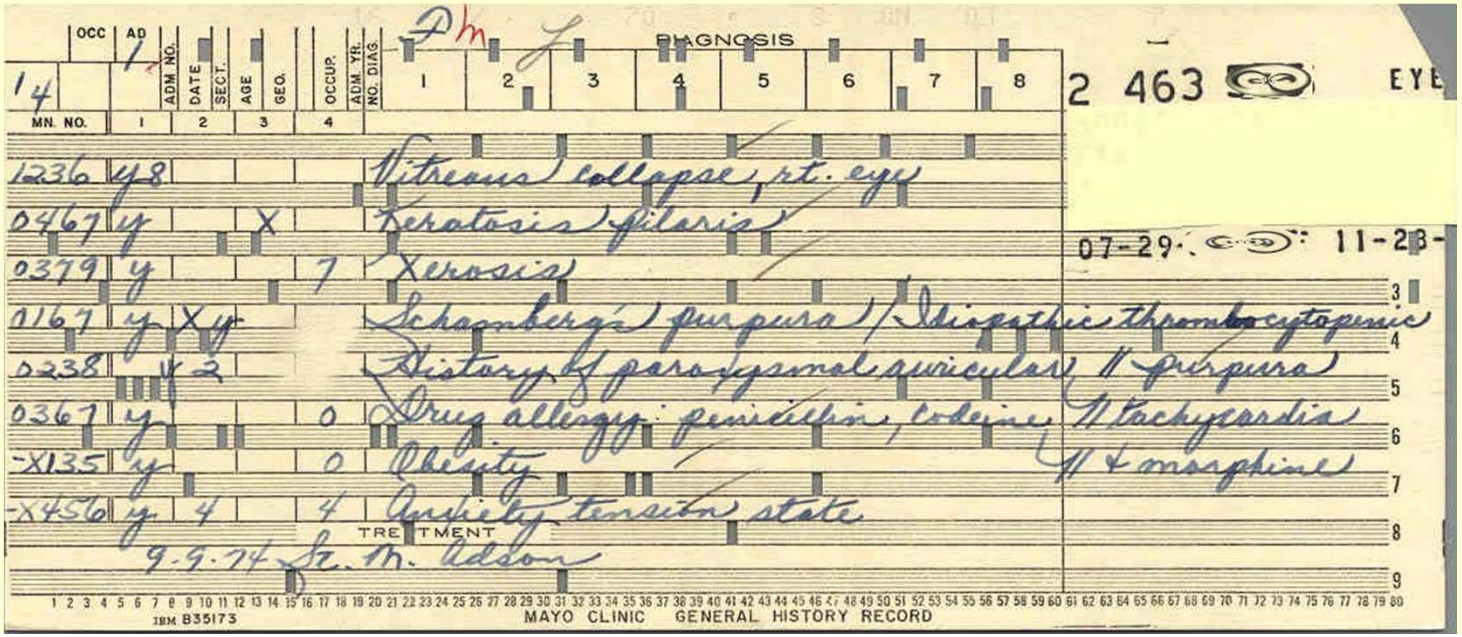
Demographics
Diagnoses
Procedures
Narratives
Laboratories
Pathology…

High-Volume Data Storage

© Mayo Clinic College of Medicine 2008

Semantic Organization





U.S. Department of Health & Human Services

http://www.hhs.gov/

Office of the National Coordinator for Health Information Technology (ONC)

Program Official: Wil Yu **http://healthit.hhs.gov**

AREA 1

University of Illinois at Urbana-Champaign

(#10510624)

Security of Health IT

PI: Carl Gunter, PhD

http://sharps.org

AREA 2

The University of Texas Health Science Center at Houston (#10510592)

Patient-Centered Cognitive Support

PI: Jiajie Zhang, PhD

http://sharpc.org

AREA 3

Harvard University (#10510924)

Healthcare Application and Network Platform Architectures

PI: Isaac Kohane, MD, PhD

Co-PI: Kenneth D. Mandl, MD, MPH

AREA 4

Mayo Clinic College of Medicine (#10510949)

Secondary Use of EHR Data

PI: Christopher Chute, MD, Dr. P.H

http://sharpn.org

SHARP: Area 4: Secondary Use of EHR Data A $15M National Consortium

- 14 academic and industry partners
- Develop tools and resources that influence and extend secondary uses of clinical data
- Cross-integrated suite of project and products
  - Clinical Data Normalization
  - Natural Language Processing (NLP)
  - Phenotyping (cohorts and eligibility)
  - Common pipeline tooling (UIMA) and scaling
  - Data Quality (metrics, missing value management)
  - Evaluation Framework (population networks)

© 2009 Mayo Clinic

Vision

- To assemble a federated informatics research community committed to open-source resources that can industrially scale to address barriers to the broad-based, facile, and ethical use of EHR data for secondary purposes
- To create, evaluate, and refine informatics artifacts that advance the capacity to efficiently leverage EHR data to improve care, generate new knowledge, and address population needs
- To make these artifacts available to the community of secondary EHR data users, manifest as open-source tools, services, and scalable software
- To partner with industry developers who can make these resources available with commercial deployment and support.

Themes & Projects

Themes			Projects	Players
Data Normalization	Phenotype Recognition	Data Quality and Evaluation Frameworks	Clinical Data Normalization	IBM, Mayo, Utah, Agilex
			Natural Language Processing (NLP)	Harvard, Group Health, IBM, Utah ,Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado
			High-Throughput Phenotyping	CDISC, Centerphase, Mayo, Utah
			UIMA and Scaling Capacity	IBM, Mayo
			Data Quality	Mayo, Utah
			Evaluation Framework	Agilex, MN HIE, Mayo, Utah

The bookends - Projects 1&6 Data Normalization & Evaluation

Christopher G. Chute

Stan Huff, Intermountain Healthcare

Overview

- Build generalizable data normalization pipeline
- Establish a globally available resource for health terminologies and value sets
- Establish and expand modular library of normalization algorithms
- Iteratively test normalization pipelines, including NLP where appropriate, against normalized forms, and tabulate discordance.
- Use cohort identification algorithms in both EMR data and EDW data. (normalize against CEMs)

Intermountain Healthcare

GE/Intermountain Clinical Element Model Search

View License Agreement | Site Information | Download All XML Models | Model Value Sets | View Current Model XML Source

Contact Us

Model Search | Model List | Model Detail

BloodPressurePanel

- SystolicBloodPressureMeas
  - PQ
  - MethodDevice
  - BodyLocationPrecoord
  - BodyPosition
  - AbnormalInterpretation
  - DeltaFlag
  - ReferenceRangeNar
  - RelativeTemporalContext
- DiastolicBloodPressureMeas
- MeanArterialPressureMeas
- MethodDevice
- BodyLocationPrecoord
- BodyPosition
- RelativeTemporalContext
- PatientPrecondition
- Subject
- Observed
- ReportedReceived
- Verified

Description / Status:

Name:	BloodPressurePanel
Definition:	BloodPressurePanel is an Associated CEM Panel that groups a systolic blood pressure, diastolic blood pressure, and mean arterial pressure all obtained at the same time.
Status:	proposed

Details | XML View

RAW XML

```
<cetype kind="panel" name="BloodPressurePanel" xmlns="">
<key code="BloodPressurePanel_KEY_ECID" />
<item card="0-1" name="systolicBloodPressureMeas" type="SystolicBloodPressureMeas" />
<item card="0-1" name="diastolicBloodPressureMeas"
type="DiastolicBloodPressureMeas" />
<item card="0-1" name="meanArterialPressureMeas" type="MeanArterialPressureMeas" />
<qual card="0-1" name="methodDevice" type="MethodDevice" />
<qual card="0-1" name="bodyLocationPrecoord" type="BodyLocationPrecoord" />
<qual card="0-1" name="bodyPosition" type="BodyPosition" />
<qual card="0-M" name="relativeTemporalContext" type="RelativeTemporalContext" />
<qual card="0-M" name="patientPrecondition" type="PatientPrecondition" />
<mod card="0-1" name="subject" type="Subject" />
```

© Mayo Clinic College of Medicine 2010

Project 2
Clinical Natural Language Processing (cNLP)

Dr. Guergana Savova
Harvard Childrens' Hospital

J Am Med Inform Assoc. 2010 Sep-Oct;17(5):507-13.

Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications

Guergana K Savova,[1] James J Masanz,[1] Philip V Ogren,[2] Jiaping Zheng,[1] Sunghwan Sohn,[1] Karin C Kipper-Schuler,[1] Christopher G Chute[1]

► Additional tables and appendices are published online only. To view these files please visit the journal online (http://jamia.bmj.com).

[1]Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota, USA
[2]Computer Science Department, University of Colorado, Denver, Colorado, USA

Correspondence to
Guergana Savova, Children's Hospital Informatics Program, Children's Hospital Boston and Harvard Medical School, 300 Longwood Avenue, Enders 138, Boston, MA 02115, USA; guergana.savova@childrens.harvard.edu

ABSTRACT

We aim to build and evaluate an open-source natural language processing system for information extraction from electronic medical record clinical free-text. We describe and evaluate our system, the clinical Text Analysis and Knowledge Extraction System (cTAKES), released open-source at http://www.ohnlp.org. The cTAKES builds on existing open-source technologies—the Unstructured Information Management Architecture framework and OpenNLP natural language processing toolkit. Its components, specifically trained for the clinical domain, create rich linguistic and semantic annotations. Performance of individual components: sentence boundary detector accuracy=0.949; tokenizer accuracy=0.949; part-of-speech tagger accuracy=0.936; shallow parser F-score=0.924; named entity recognizer and system-level evaluation F-score=0.715 for exact and 0.824 for overlapping spans, and accuracy for concept mapping,

NLP system designed to process and extract semantically viable information to support the heterogeneous clinical research domain and to be sufficiently scalable and robust to meet the rigors of a clinical research production environment. This paper describes and evaluates our system—the clinical Text Analysis and Knowledge Extraction System (cTAKES).

BACKGROUND

The clinical narrative has unique characteristics that differentiate it from scientific biomedical literature and the general domain, requiring a focused effort around methodologies within the clinical NLP field.[2] Columbia University's proprietary Medical Language Extraction and Encoding System (MedLEE)[3] was designed to process radiology reports, later extended to other domains,[4] and tested for transferability to another institu-

Project 3
High throughput Phenotyping (HTP)

Dr. Jyoti Pathak
Mayo Clinic

Phenotypes – Overload of term

- Originally for research cohorts from EMRs
- Obvious extension to clinical trial eligibility
- Note relevance to quality metrics
  - Numerator and denominator constitute phenotypes
- Clinical decision support
  - Trigger criteria

© Mayo Clinic College of Medicine 2010

Project 4
Infrastructure & Scalability

Calvin Beebe, Mayo Clinic, IT
Marshal Schor, IBM, Watson Labs
Jeff Ferraro, Group Health, Seattle

UIMA is the "Jeopardy" engine framework

- http://www.jeopardy.com/minisites/watson/

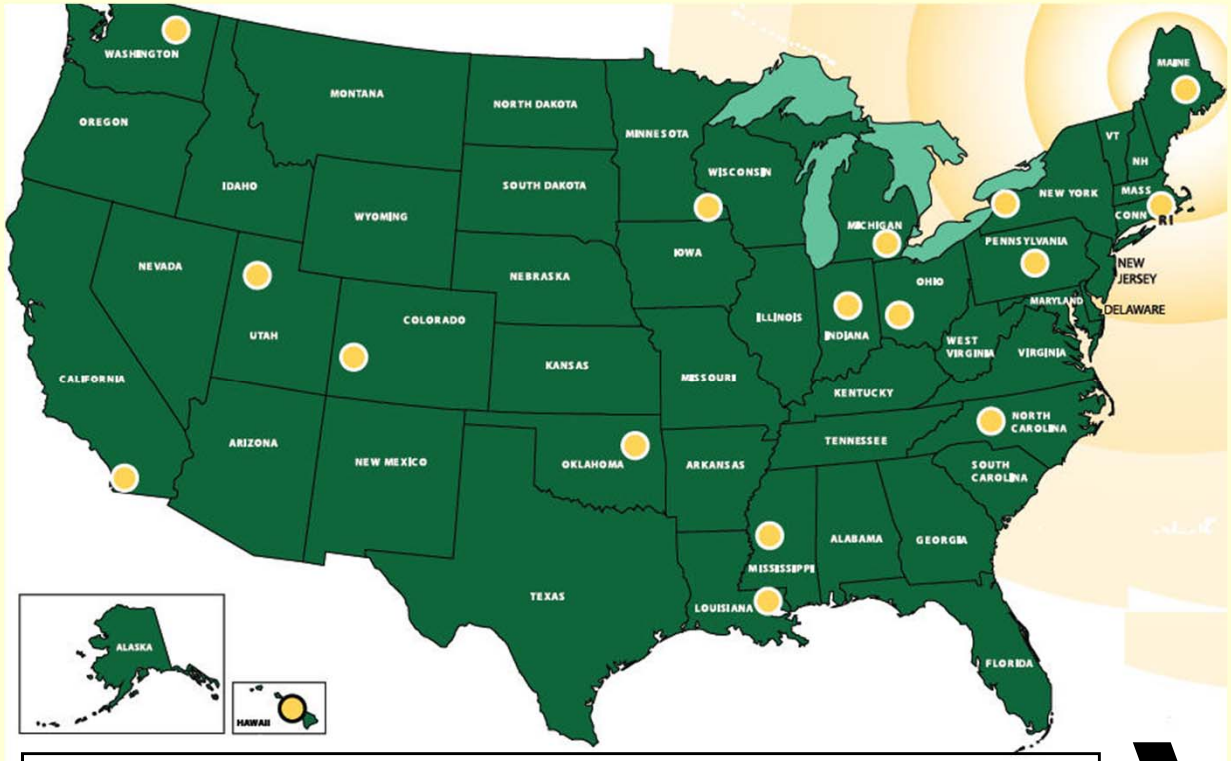
© 2009 Mayo Clinic

UIMA at Mayo

- Mayo was the Alpha user of UIMA in 2003
  - Foundation of Natural Language Processing
- Mayo convinced IBM to make UIMA open-source
- Continued major use of UIMA within informatics
- Anchoring framework for SHARP
- Target of Regenstrief HOSS transform
- Proposed as translation engine for Beacon

© 2009 Mayo Clinic

17 Beacon Communities



Extend advanced health IT & exchange infrastructure

Leverage data to inform specific delivery system & payment strategies

Demonstrate a vision of the future where:

- **Hospitals, clinicians, & patients are meaningful users of health IT**
- **Communities achieve measurable & sustainable improvements in health care quality, safety, efficiency, and population health**

Beacon: $12.3M Healthcare improvement through standards-based health information exchange

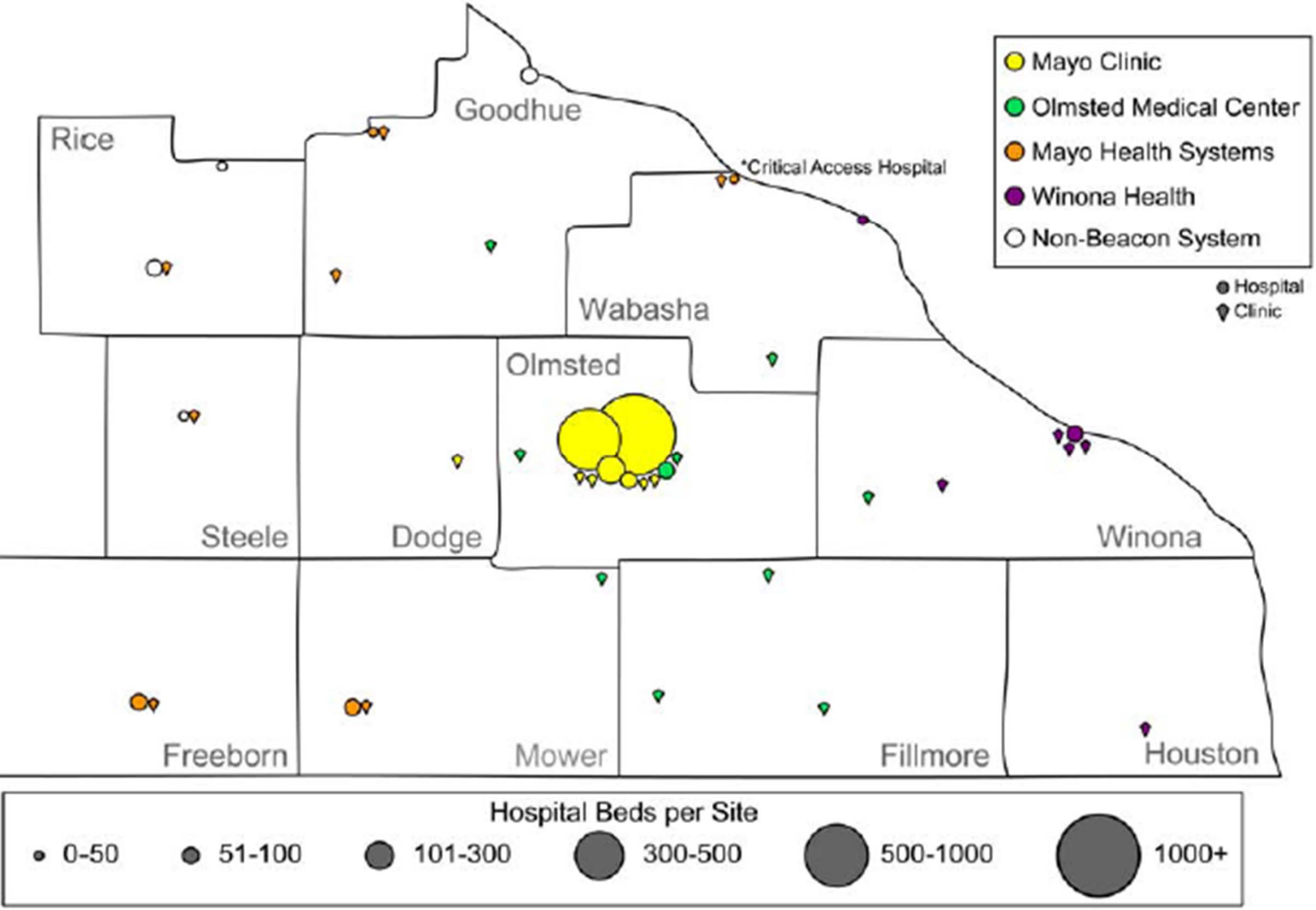
- 11 SE MN counties; all public health depts, Mayo, Winona Health System, Olmsted Medical Center, MN Health Information Exchange, Indian Health Service, VA, Federal Med Center
- Demonstrate efficacy in childhood asthma & DM
  - Implement standards-based data exchange
  - National data exchange methods and monitoring
  - Comparative effectiveness research
  - Invoke 6-sigma and LEAN quality improvement
  - Disease metrics, utilization of ER and hospital

© 2009 Mayo Clinic

SE Minnesota Vision:

- Establish SE MN as "the premier population-based laboratory for healthcare delivery, best-practices discovery, reimbursement modeling, clinical quality optimization, and translational research in the country."

© 2009 Mayo Clinic


Mayo Clinic
Olmsted Medical Center
Mayo Health Systems
Winona Health
Non-Beacon System
Hospital
Clinic
*Critical Access Hospital
Rice
Goodhue
Wabasha
Olmsted
Steele
Dodge
Winona
Freeborn
Mower
Fillmore
Houston
Hospital Beds per Site
0-50
51-100
101-300
300-500
500-1000
1000+

SHARP and Beacon Synergies

- High-throughput data normalization and phenotyping (SHARP)
- Applied to population laboratory (Beacon)
- Validate on consented sub-samples
- Potential to include ALL patients in population area – regardless of provider

© 2009 Mayo Clinic

SHARP Contributions to Preservation EMR Data Normalization

- Syntactic: Transform idiosyncratic data and information into canonical standard form
  - Meaningful Use Standards
  - Clinical Element Models
- Semantic: Annotate named entities
  - Documented and versioned value sets
  - Derived from or mapped to standard terminologies
- Consequence of preparation for secondary use enhances interpretability into the future

© 2009 Mayo Clinic

SHARP Area 4: More information…

Log in / cre

page | discussion | view source | history

navigation
- Main Page
- News
- Community portal
- Help

search

Go | Search

toolbox
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

Main Page

Strategic Health IT Advanced Research Projects (SHARP)

Research Focus Area: 4. Secondary Use of EHR Data

http://informatics.mayo.edu/sharp

http://sharpn.org

This is a collaborative project management platform.

Proposal

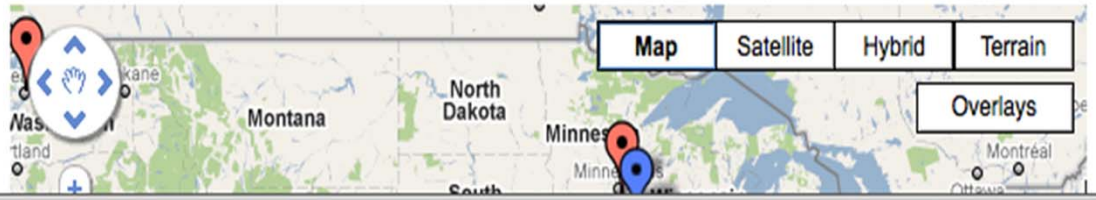
We propose research that will generate a framework of open-source services that can be dynamically configured to transform EHR data into standards-conforming, comparable infor
suitable for large-scale analyses, inferencing, and integration of disparate health data. We will read more.

SHARP Program Organization

Contacts

- **Principal Investigator:** Christopher G. Chute, MD, DrPH, 507-284-5506, chute@mayo.edu
- **Program Manager:** Lacey Hart, MBA, PMP, 507-293-3780, hart.lacey@mayo.edu
- **Media Relations:** Robert Nellis, 507-284-5005, newsbureau@mayo.edu
- **ONC Program Official:** Wil Yu, 202-690-7151, Wil.Yu@hhs.gov
- Program Advisory Committee

Collaborators



Southeast Minnesota Beacon Program

Eleven Minnesota counties working to improve health care and health delivery systems.

Home Blog Community Engagement Diabetes Partners Asthma News

Welcome to Beacon

Beacon is a community-based program to spotlight a variety of "best practice" approaches to improving health and health care delivery in the United States. Funded by the U.S. Department of Health and Human Services, through the Office of the National Coordinator for Health Information Technology, the Beacon Communities are a series of medical practice and research coalitions focusing on specific health conditions in their areas and utilizing and developing efficient systems based on their foundational expertise.

Overall, each Beacon program seeks to:

Search

Recent Posts
- Beacon is beginning

Archives
Select Month

Categories
Select Category

Project Team Wiki
- Project Team Wiki

© 2009 Mayo Clinic