Strategic Health IT Advanced Research Projects (SHARP)
Area 4: Secondary Use of EHR Data

SHARPfest – Washington
June 2-3, 2010

PI: Christopher G Chute, MD DrPH
Collaborations

- Agilex Technologies
- CDISC (Clinical Data Interchange Standards Consortium)
- Centerphase Solutions
- Deloitte
- Group Health, Seattle
- IBM Watson Research Labs
- University of Utah

- Harvard Univ. & i2b2
- Intermountain Healthcare
- Mayo Clinic
- Minnesota HIE (MNHIE)
- MIT and i2b2
- SUNY and i2b2
- University of Pittsburgh
- University of Colorado
Project Advisory Committee

Suzanne Bakken, RN DNSc, Columbia University
C. David Hardison, PhD, VP SAIC
Barbara A. Koenig, PhD, Bioethics, Mayo Clinic
Issac Kohane, MD PhD, i2b2 Director, Harvard
Marty LaVenture, PhD MPH, Minnesota Department of Health
Dan Masys, MD, Chair, Biomedical Informatics, Vanderbilt University
Mark A. Musen, MD PhD, Division Head BMIR, Stanford University
Robert A. Rizza, MD, Executive Dean for Research, Mayo Clinic
Nina Schwenk, MD, Vice Chair Board of Governors, Mayo Clinic
Kent A. Spackman, MD PhD, Chief Terminologist, IHTSDO
Tevfik Bedirhan Üstün, MD, Coordinator Classifications, WHO
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

<table>
<thead>
<tr>
<th>Themes</th>
<th>Projects</th>
<th>Players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Normalization</td>
<td>Clinical Data Normalization</td>
<td>IBM, Mayo, Utah, Agilex</td>
</tr>
<tr>
<td>Phenotype Recognition</td>
<td>Natural Language Processing (NLP)</td>
<td>Harvard, Group Health, IBM, Utah, Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado</td>
</tr>
<tr>
<td>Data Quality and Evaluation Framework</td>
<td>High-Throughput Phenotyping</td>
<td>CDISC, Centerphase, Mayo, Utah</td>
</tr>
<tr>
<td>Data Quality</td>
<td>UIMA and Scaling Capacity</td>
<td>IBM, Mayo</td>
</tr>
<tr>
<td>Evaluation Framework</td>
<td>Data Quality</td>
<td>Mayo, Utah</td>
</tr>
<tr>
<td></td>
<td>Evaluation Framework</td>
<td>Agilex, MN HIE, Mayo, Utah</td>
</tr>
</tbody>
</table>
Project 1 - Clinical Data Normalization
CG Chute, MD DrPH

• Build generalizable data normalization pipeline
• Semantic normalization annotators involving LexEVS
• Establish a globally available resource for health terminologies and value sets
• Establish and expand modular library of normalization algorithms
Project 2: Clinical Natural Language Processing (cNLP); GK Savova, PhD

• Overarching goal
  • High-throughput phenotype extraction from clinical free text based on standards and the principle of interoperability

• Focus
  • Information extraction (IE): transformation of unstructured text into structured representations
  • Merging clinical data extracted from free text with structured data
Project 3: High-Throughput Phenotyping

Jyoti Pathak, PhD

- Develop portable phenotype algorithms
  - Administrative Data; Labs; Medications
  - Build on NHGRI eMERGE experience

- Phenotyping logic specification

- Applications of phenotype characterization

- Expansion of Cohort Amplification methods
Project 4 - UIMA exploitation
Marshall Schor – IBM Research

• Use UIMA as a unifying framework, leveraging ecosystem
  • Work with team leads to identify “fit” (or not) of UIMA into subprojects
    • Phenotyping and Data Quality, especially
• Support UIMA and UIMA-AS use
  • Consult on pipe line design / architectures / configuration
• Support scaling, capacity flexibility
  • Develop and deploy virtual machine images that can dynamically scale in cloud computing environments
  • Develop integration / deployment tooling with goal of simplicity
    • Enabling widespread adoption of POC
Project 5 - Data Quality
Kent Bailey, PhD

• Refine metrics for data consistency

• Deploy methods for missing or conflicting data resolution

• Integrate methods into UIMA pipelines

• Refine and enhance methods
Project 6 - Real-world evaluation framework
Dr. Huff

• Iteratively test normalization pipelines, including NLP where appropriate, against normalized forms, and tabulate discordance.
  • Normalize retrospective data from the EMRs and compare it to normalized data that already exists in our data warehouses (Mayo Enterprise Data Trust, Intermountain).

• Use cohort identification algorithms in both EMR data and EDW data.
  • Normalize the data against CEMs.
Potential NHIN Incorporation

NHIN

EMR

UIMA

Canonical Models

Decision Support

CER HTA QI CDS

Terminology Services (including CEMs)

EDW Staging

ETL + Rules

Analytic Health Repository

Decision Support

CER HTA QI CDS

NLP
Area 4: More information…

Main Page

Strategic Health IT Advanced Research Projects (SHARP)

Research Focus Area: 4. Secondary Use of EHR Data

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 information 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 Neill, 507-284-5005, newsbureau@mayo.edu
- **ONC Program Official:** Wil Yu, 202-590-7151, Wil.Yu@hhs.gov
- **Program Advisory Committee**

Collaborators

http://informatics.mayo.edu/sharp
http://sharpmn.org