Secondary Use of EMR Data
View from SHARPn
AMIA Health Policy, 12 Dec 2012

Christopher G. Chute, MD DrPH,
Professor, Biomedical Informatics, Mayo Clinic
Chair, ISO TC215 on Health Informatics
Chair, International Classification of Disease, WHO
Declarations

☐ No real or apparent financial conflicts of interest
  – All products are open-source

☐ Comments represent beliefs of the author
Secondary Use

Analyses or interpretation of clinical data across multiple patients

- Clinical Quality Improvement
- Comparative Effectiveness Analyses
- Outcomes Research
- Best Evidence Discovery
- Technology Assessment
- Data-driven Clinical Decision Support
From Practice-based Evidence to Evidence-based Practice

Data
- Clinical Databases
- Registries et al.

Inference
- Standards
  - Comparable and Consistent
  - Medical Knowledge
  - Vocabularies & Terminologies

Decision Support
- Expert Systems
- Clinical Guidelines

Knowledge Management

"Secondary Use"

© 2012 Mayo Clinic
Comparable and Consistent Data

- Inferencing from data to information requires sorting information into categories
  - Statistical bins
  - Machine learning features
- Accurate and reproducible categorization depends upon semantic consistency
- Semantic consistency is the vocabulary problem
  - Almost always manifest as the “value set” problem

© 2012 Mayo Clinic
The Challenge

- Most clinical data in the United States is heterogeneous – non-standard
  - Within Institutions
  - Between Institutions

- Meaningful Use is mitigating, but has not yet “solved” the problem
  - Achieving standardization in Meaningful Use is sometimes minimized
SHARP Area 4: Secondary Use of EHR Data

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

- Harvard Univ.
- Intermountain Healthcare
- Mayo Clinic
- Mirth Corporation, Inc.
- MIT
- MITRE Corp.
- Regenstrief Institute, Inc.
- SUNY
- University of Colorado
Cross-integrated, *open-source*, suite of projects and products

<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, Regenstrief</td>
</tr>
<tr>
<td>Phenotype Recognition</td>
<td>Natural Language Processing</td>
<td>Harvard, Group Health, IBM, Utah, Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado, MITRE</td>
</tr>
<tr>
<td>Data Quality &amp; Evaluation Frameworks</td>
<td>High-Throughput Phenotyping</td>
<td>CDISC, Centerphase, Mayo, Utah</td>
</tr>
<tr>
<td></td>
<td>UIMA &amp; Scaling Capacity</td>
<td>IBM, Mayo, Agilex, Mirth</td>
</tr>
<tr>
<td></td>
<td>Data Quality</td>
<td>Mayo, Utah</td>
</tr>
<tr>
<td></td>
<td>Evaluation Framework</td>
<td>Agilex, Mayo, Utah</td>
</tr>
</tbody>
</table>
SHARPn Tools

- Library or suite of open-source tools
  - Apache 2.0, Commercial friendly
- Positioned as Middle-ware
- Intended to work on EMR “messages”
  - Defined by Meaningful Use standards
  - HL7 V2.51, cCDA, CCD, etc.
  - Clinical text, narratives, reports

© 2012 Mayo Clinic
Modes of Normalization

- Generally true for both *structured* and *unstructured* data
- Syntactic transformation
  - Clean up message formats
  - HL7 V2, CCD/CDA, tabular data, etc
  - Emulate Regenstrief HOSS pipeline
- Semantic normalization
  - Typically vocabulary mapping
Transformation Target?

- Normalization begs a “normal form”
- Extant national and international standards do not fully specify
  - Focus on HIE or internal messaging
  - Canonical data representation wanting
  - Require fully machine manageable data
Clinical Data Normalization

- Data Normalization
  - Comparable and consistent data is foundational to secondary use

- Clinical Data Models – Clinical Element Models (CEMS)
  - Basis for retaining computable meaning when data is exchanged between heterogeneous computer systems.
  - Basis for shared computable meaning when clinical data is referenced in decision support logic.
**BloodPressurePanel**

- SystolicBloodPressureMeas
- DiastolicBloodPressureMeas
- MeanArterialPressureMeas
- MethodDevice
- BodyLocationPrecoord
- BodyPosition
- AbnormalInterpretation
- DeltaFlag
- ReferenceRangeNar
- RelativeTemporalContext

**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

**RAW XML**

```xml
<ctype 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"/>
```
Data Element Harmonization

http://informatics.mayo.edu/CIMI/

- Stan Huff – CIMI
  - Clinical Information Model Initiative
- NHS Clinical Statement
- CEN TC251/OpenEHR Archetypes
- HL7 Templates
- ISO TC215 Detailed Clinical Models
- CDISC Common Clinical Elements
- Intermountain/GE CEMs

© 2012 Mayo Clinic
Core CEMs

- Recognize that use-case specific work-flow enters into CEM-like objects
  - Clinical EHR implementation
  - CLIA or FDA regulatory overhead
- Secondary Use tends to focus on data
- Create “core” CEMs
  - Labs, Rxs, Dxs, Pxs, demographics
That Semantic Bit…

- Canonical semantics reduce to Value-set Binding to CEM objects
- Value-sets drawn from “standard” vocabularies
  - SNOMED-CT and ICD, LOINC, RxNorm
- Common Terminology Services (CTS2)
- NLM National Value-set Center
  - CTS2 Value-set services
Normalization Pipelines

- Input heterogeneous clinical data
  - HL7, cCDA/CCD, structured feeds

- Output Normalized CEMs
  - Create logical structures within UIMA CAS

- Serialize to a persistence layer
  - SQL, RDF, “PCAST like”, XML

- Robust Prototypes now posted
  - Early version production Q3 2012
NLP Deliverables and Tools

http://informatics.mayo.edu/sharp/index.php/Tools

- cTAKES Releases
  - Smoking Status Classifier
  - Medication Annotator
  - cTAKES Side Effects module
  - Modules for relation extraction

- Integrated cTAKES (icTAKES)
  - an effort to improve the usability of cTAKES for end users

- NLP evaluation workbench
  - the dissemination of an NLP algorithm requires performance benchmarking. The evaluation workbench allows NLP investigators and developers to compare and evaluate various NLP algorithms.

- SHARPn NLP Common Type
  - SHARPn NLP Common Type System is an effort for defining common NLP types used in SHARPn; UIMA framework.

© 2012 Mayo Clinic
High-Throughput Phenotyping

- Phenotype - a set of patient characteristics:
  - Diagnoses, Procedures
  - Demographics
  - Lab Values, Medications

- Phenotyping – overload of terms
  - Originally for research cohorts from EMRs
  - Obvious extension to clinical trial eligibility
  - Quality metric Numerators and denominators
  - Clinical decision support - Trigger criteria
EMR Phenotype Algorithms

- Typical components
  - Billing and diagnoses codes; Procedure codes
  - Labs; Medications
  - Phenotype-specific co-variates (e.g., Demographics, Vitals, Smoking Status, CASI scores)
  - Pathology; Imaging?

- Organized into inclusion and exclusion criteria

- Experience from eMERGE Electronic Medical Records and Genomics Network (http://www.gwas.net)
SHARP and Beacon Synergies

- SHARP will facilitate the mapping of comparable and consistent data into information and knowledge
- SE MN Beacon will facilitate the population-based generation of best evidence and new knowledge
- SE MN Beacon will allow the application of Health Information Technology to primary care practice
  - Informing practice with population-based data
  - Supporting practice with knowledge

© 2012 Mayo Clinic
Principles and Conclusions

☐ Comparable and Consistent Data

☐ Importance of Canonical Form
  – At source or by transformation

☐ Standards are Evolving – Rapidly

☐ Meaningful Use is Good Start
  – Continued evolution expected