SHARPn.org

Strategic Investments Towards Interoperability

Redwood MedNet, 25 July 2013

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

Strategic Health IT Advanced Research Projects (SHARP) Program

Awardee of The Office of the National Coordinator for Health Information Technology

Declarations

- No real or apparent financial conflicts of interest
  - All products are open-source
- Comments represent beliefs of the author

SHARPn.org

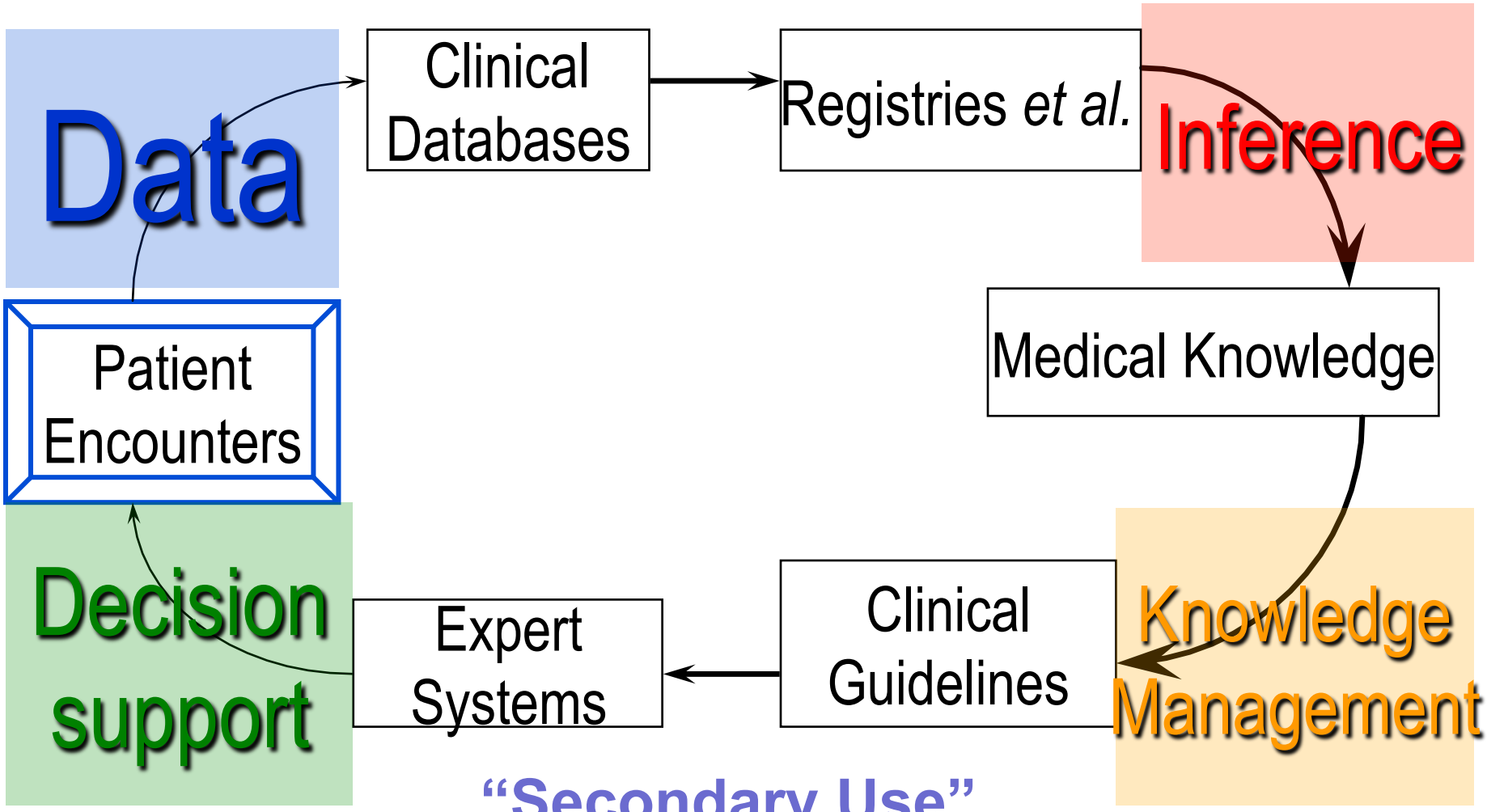
Secondary Use

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

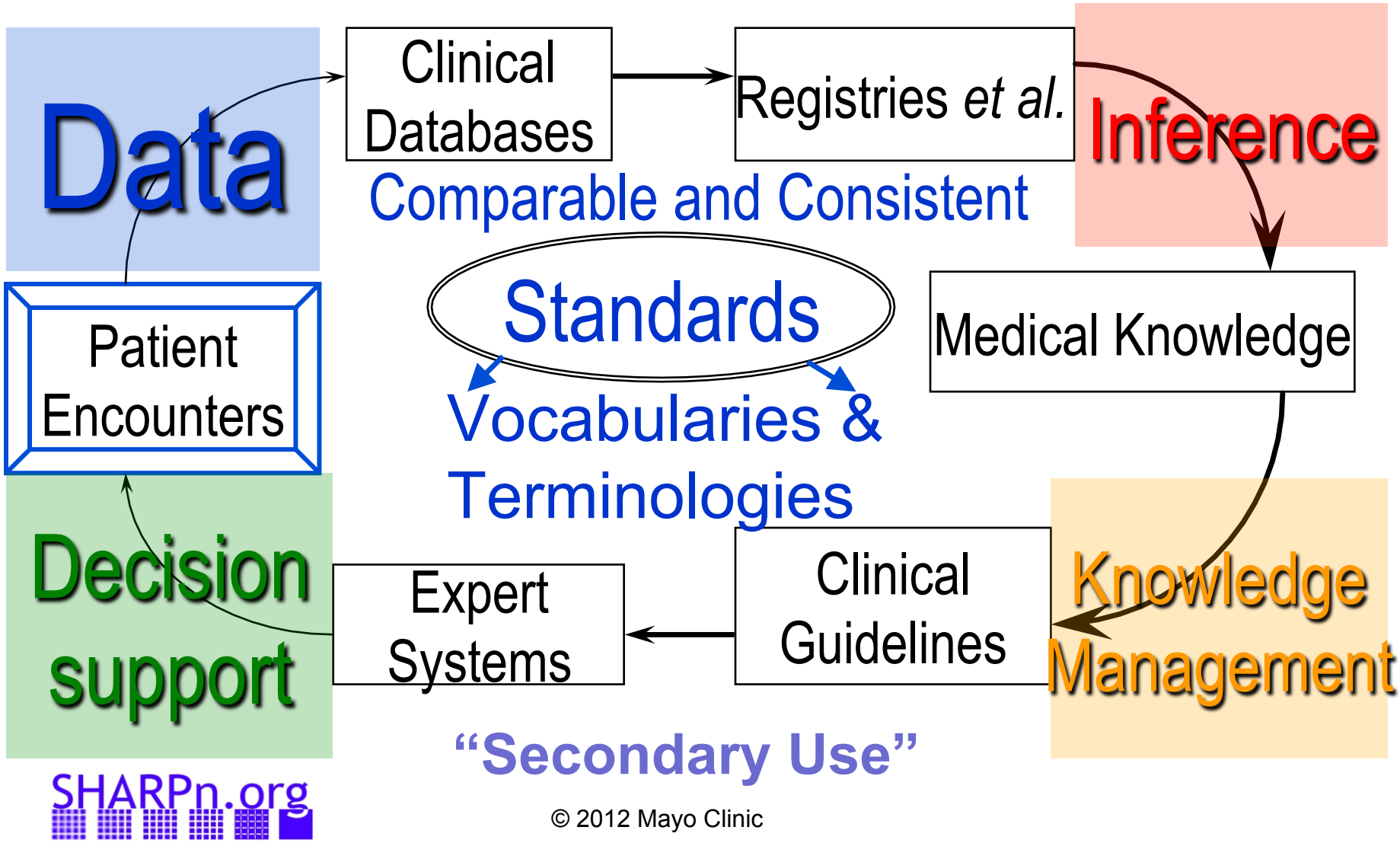
SHARPn.org

From Practice-based Evidence to Evidence-based Practice



SHARPn.org

From Practice-based Evidence to Evidence-based Practice



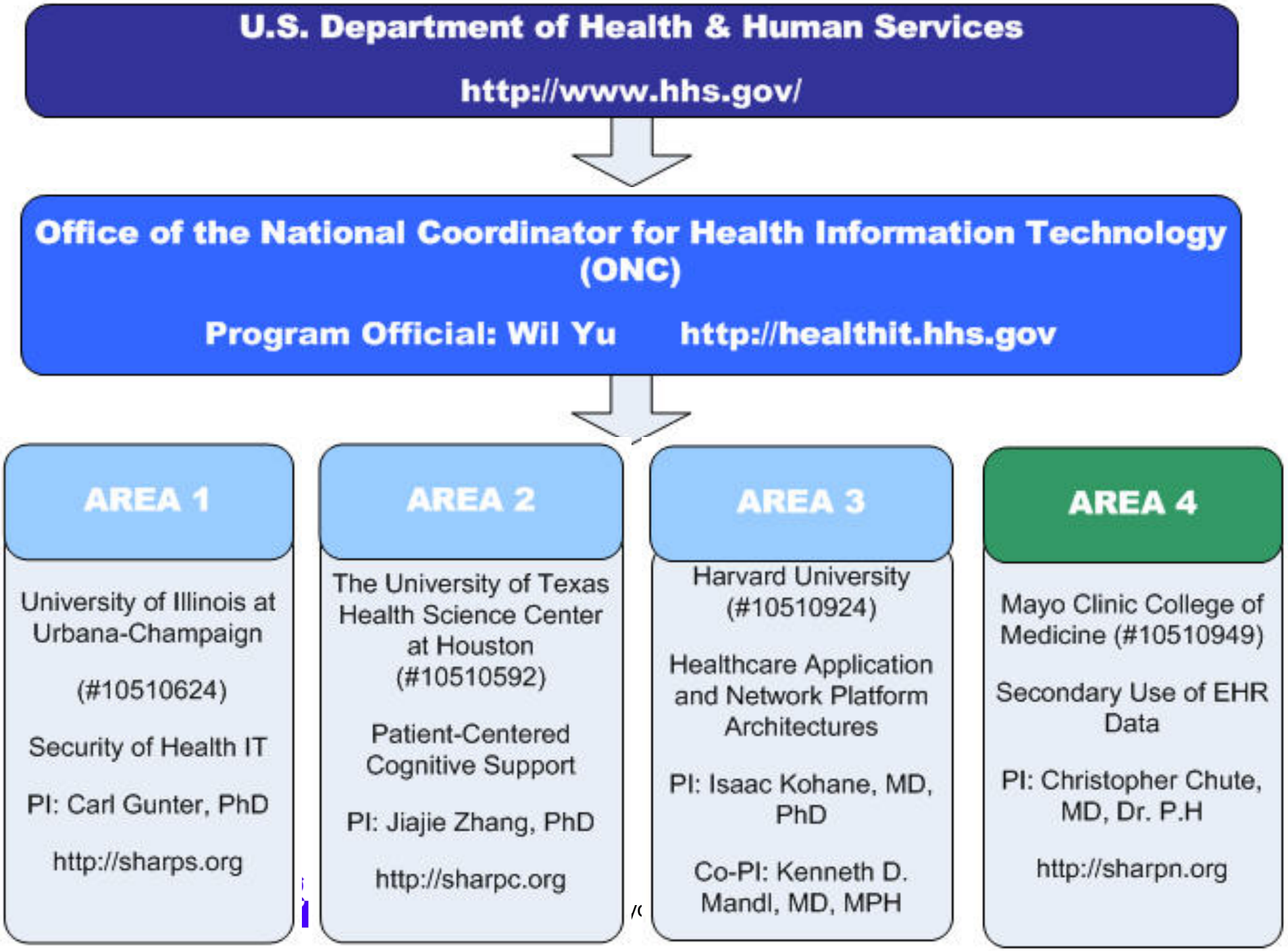
SHARPn.org

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

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


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

- 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

Themes	Projects	Players
Data Normalization, Phenotype Recognition, Data Quality & Evaluation Frameworks	Clinical Data Normalization	IBM, Mayo, Utah, Agilex, Regenstrief
	Natural Language Processing	Harvard, Group Health, IBM, Utah, Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado, MITRE
	High-Throughput Phenotyping	CDISC, Centerphase, Mayo, Utah
	UIMA & Scaling Capacity	IBM, Mayo, Agilex, Mirth
	Data Quality	Mayo, Utah
	Evaluation Framework	Agilex, Mayo, Utah

SHARPn.org

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

SHARPn.org

Modes of Normalization

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

SHARPn.org

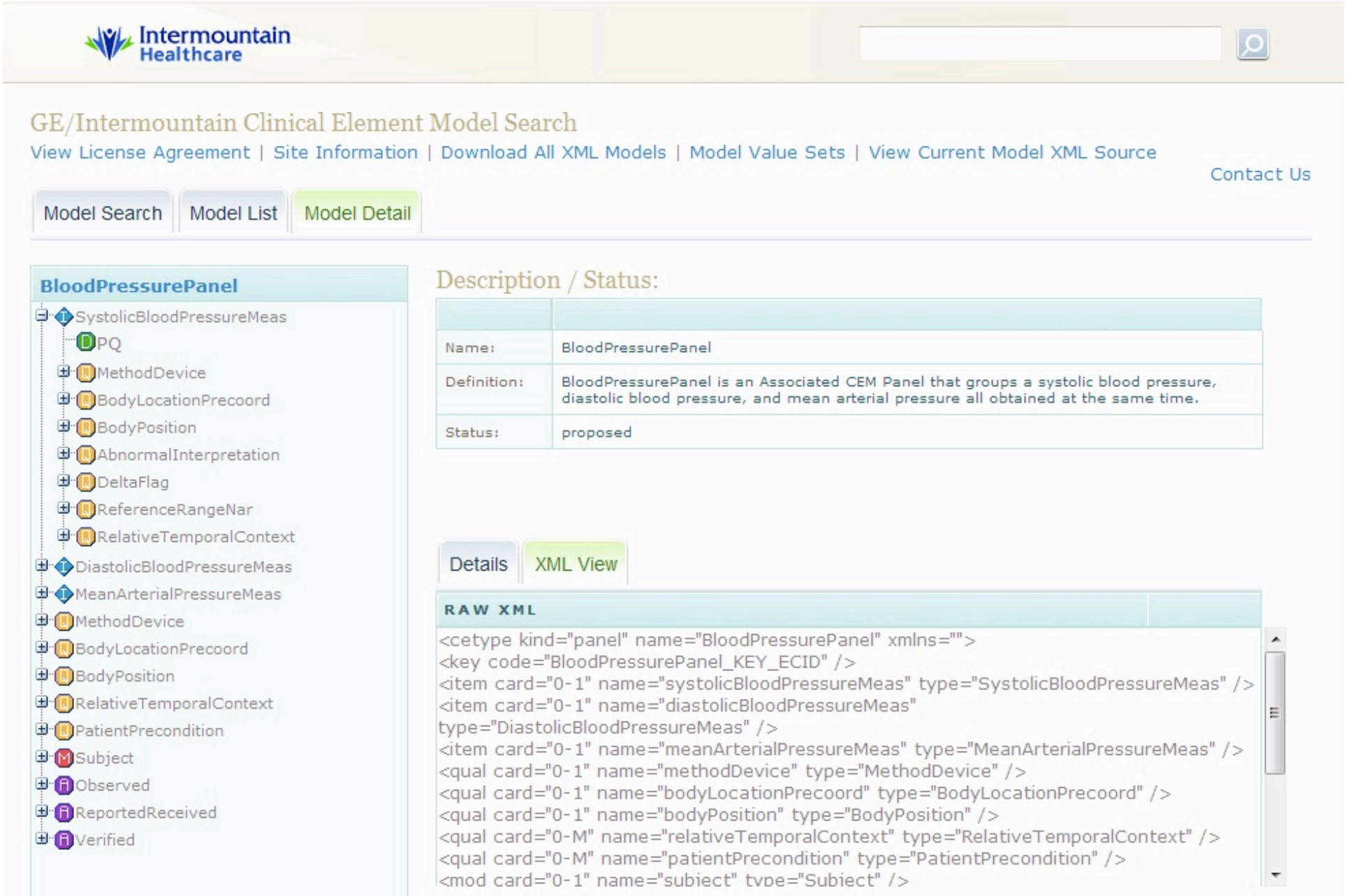
Transformation Target?

- 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.

SHARPn.org



SHARPn.org

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

SHARPn.org

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

SHARPn.org

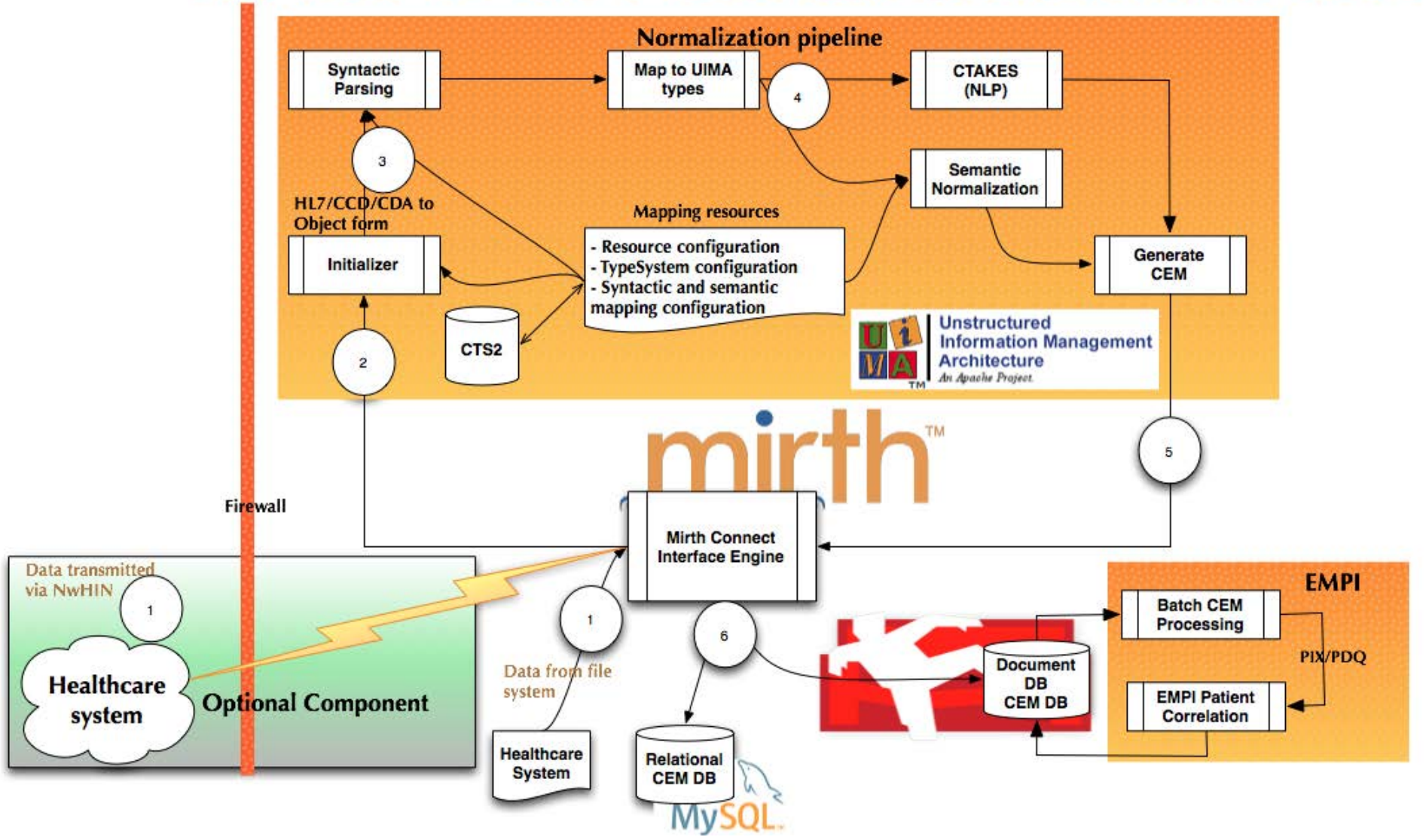
Normalization Pipelines

- 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

SHARPn.org

https://sourceforge.net/p/sharpn/datan/code/4/tree/trunk/

SHARPN Data Normalization Architecture



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.

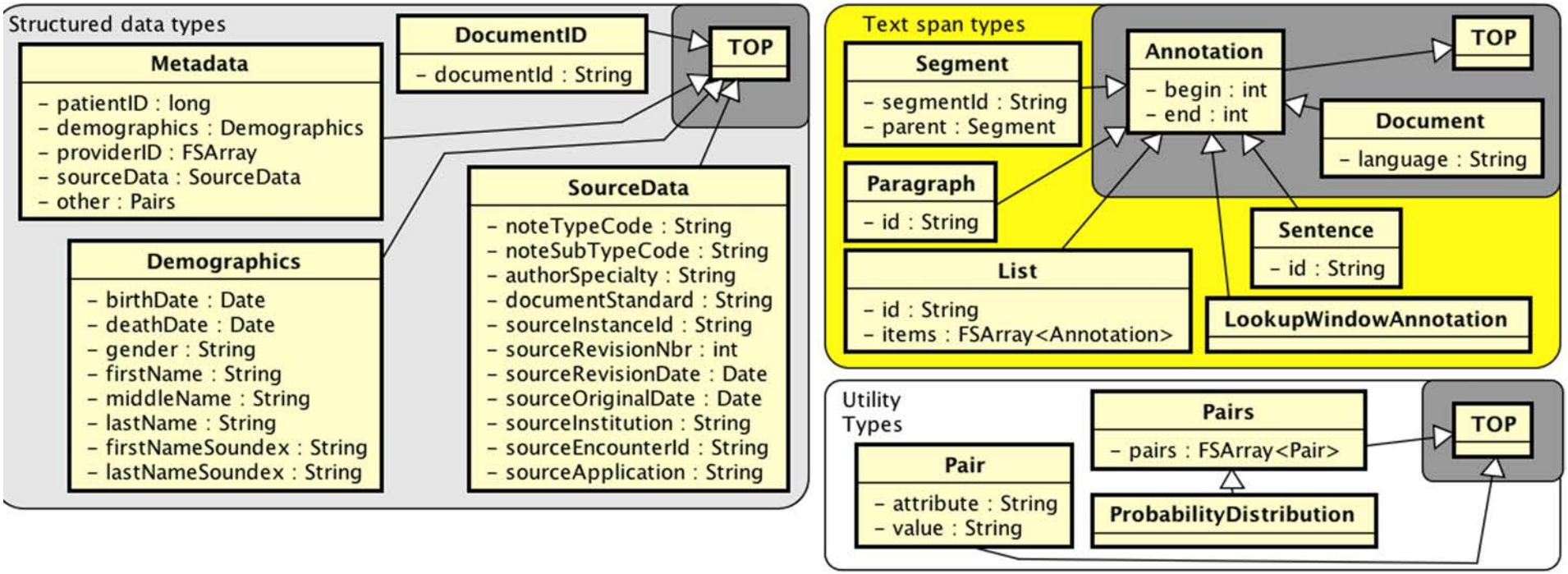
SHARPn.org

Specialized NLP Tools

- MedTagger –contextual clinical concepts extraction (certainty, status, and experiencer) – (i2b2 winner)
- MedXN – medication detection and normalization tool that associates medication mentions to clinical drugs in RxNorm
- SharpnDN – normalization clinical information across diverse data sources to common format
- MedTime – detect and normalize time expression from clinical narrative- The best system internationally
- meTAKES – an integrated NLP front-end for accessing data
  - Semantic, Indexing, Just-in-time-annotation, multi-layer language information retrieval
- Many phenotype-specific NLP tools – DILI, HF, PAD, DVT, Asthma

SHARPn.org

NLP Common Type Systems



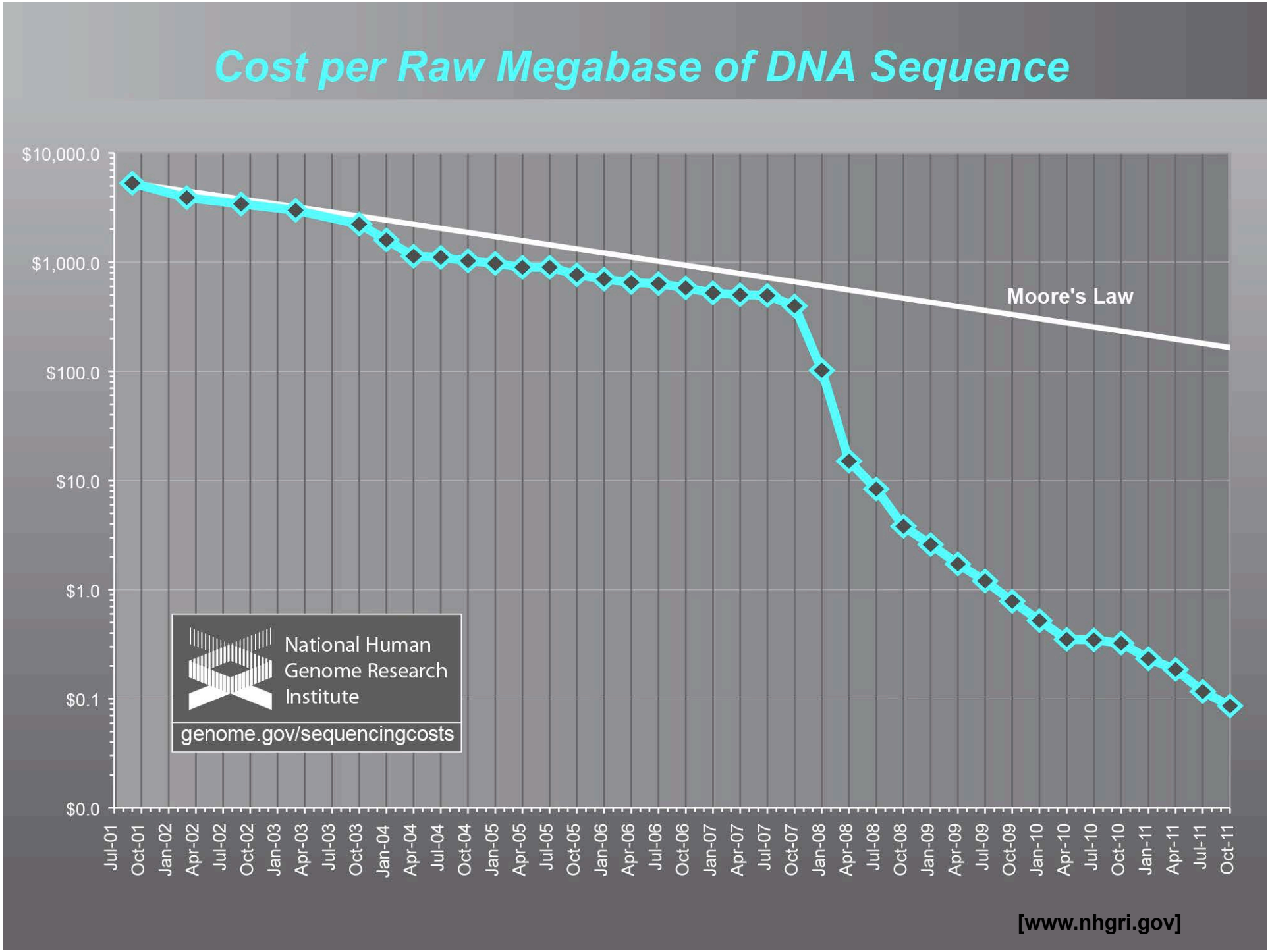
SHARPn.org

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

SHARPn.org

Cost per Raw Megabase of DNA Sequence

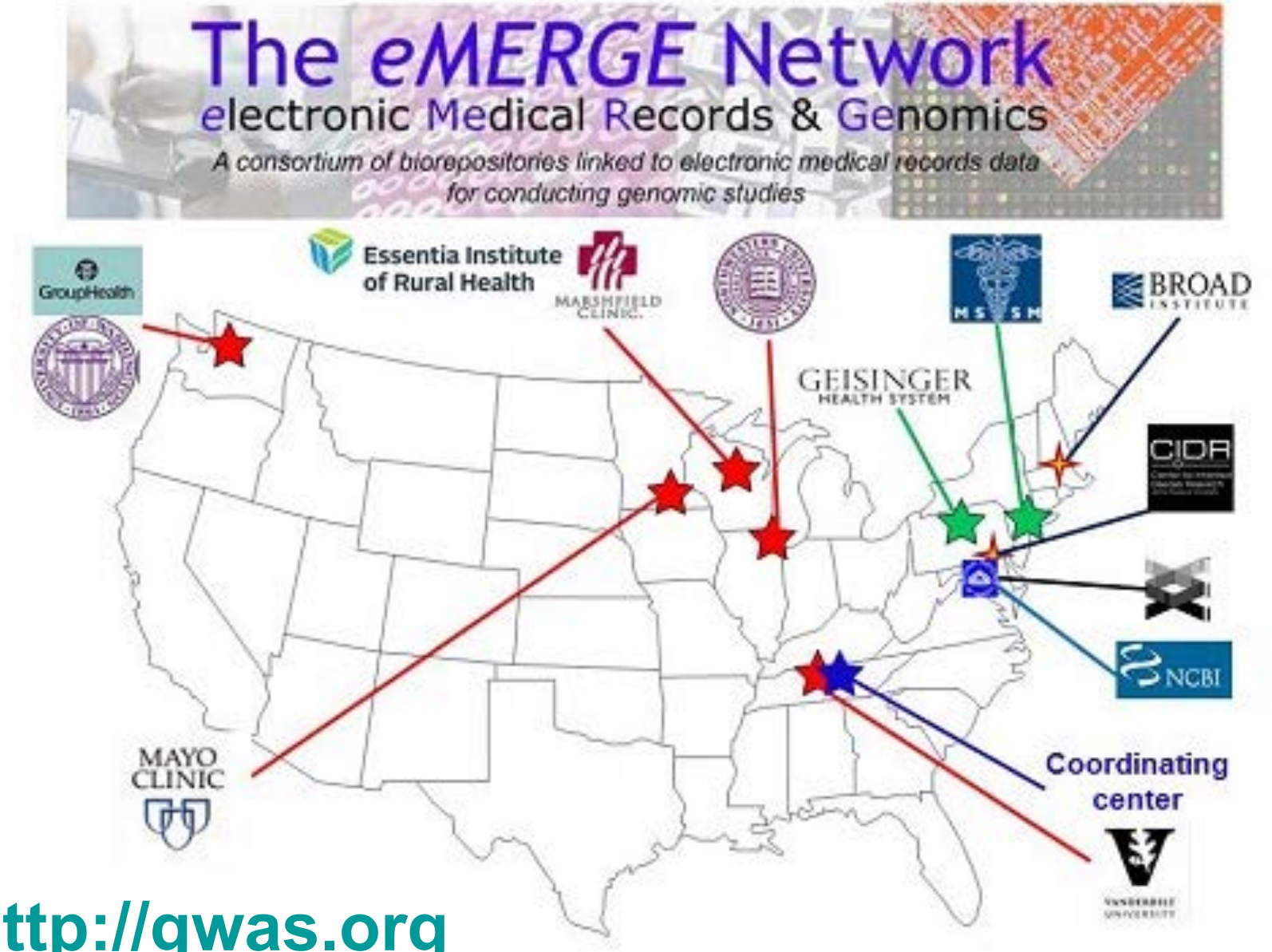


[www.nhgri.gov]

EMR Phenotype Algorithms I

- 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)

SHARPn.org

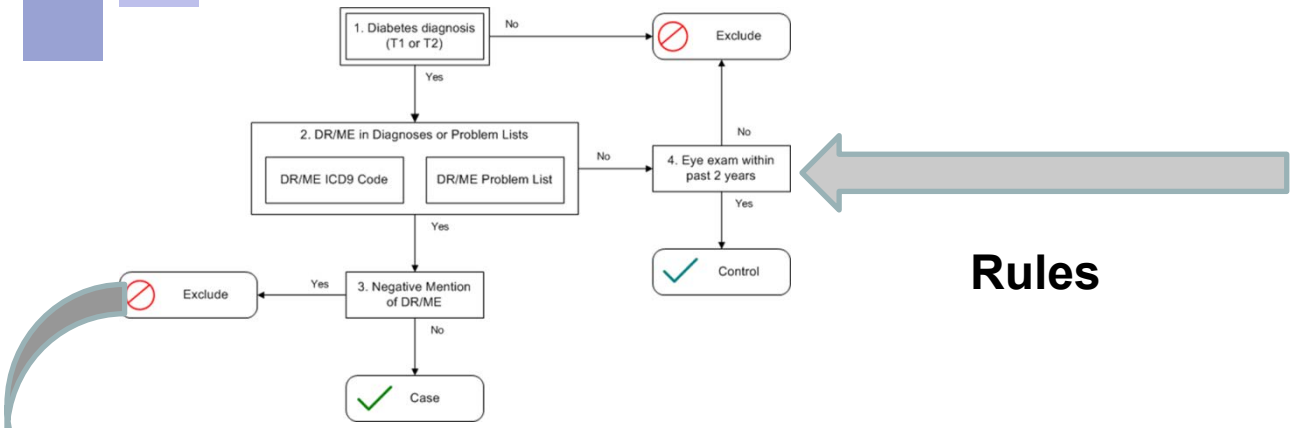


http://gwas.org

SHARPn.org

[eMERGE Network]

EHR-driven Phenotyping Algorithms II



Rules

Phenotype Algorithm

Evaluation

Visualization

Transform

Transform

Data

Mappings

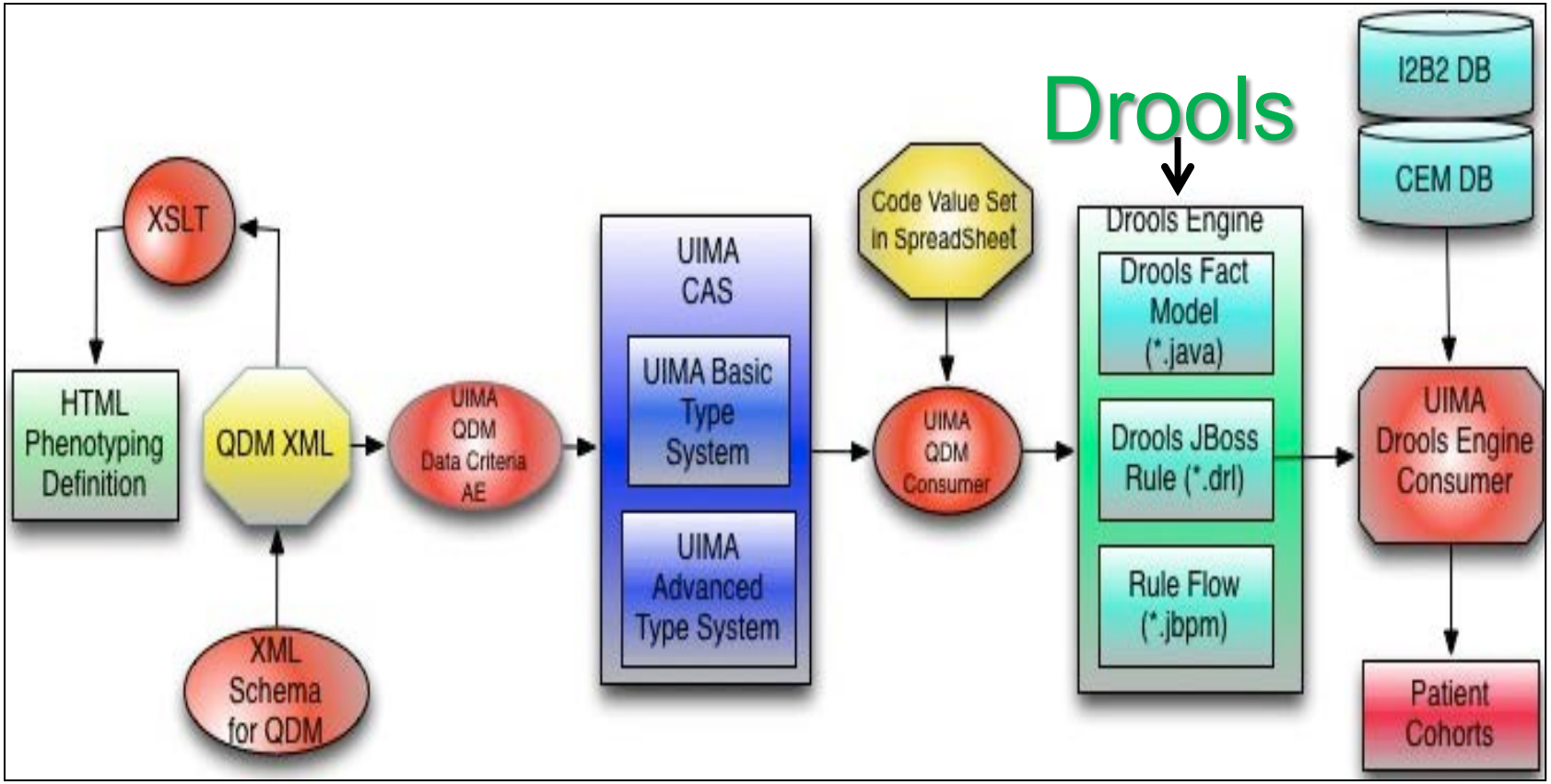
NLP, SQL

[eMERGE Network]

SHARPn.org

Modeling and Executing Electronic Health Records Driven Phenotyping Algorithms using the NQF Quality Data Model and JBoss® Drools Engine

Dingcheng Li, PhD[1], Gopu Shrestha, MS[1], Sahana Murthy, MS[1] Davide Sottara, PhD[2]
Stanley M. Huff, MD[3] Christopher G. Chute, MD, DrPH[1] Jyotishman Pathak, PhD[1]
[1]Mayo Clinic, Rochester, MN [2]University of Bologna, Italy [3]Intermountain Healthcare, Salt Lake City, UT



SHARPn.org

[Li et al., AMIA 2012; (Epub ahead of print)]

SHARPn: Secondary Use of EHR Data

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project

Susan Rea [a,*], Jyotishman Pathak [b], Guergana Savova [c], Thomas A. Oniki [d], Les Westberg [e], Calvin E. Beebe [b], Cui Tao [b], Craig G. Parker [a], Peter J. Haug [a,f], Stanley M. Huff [d,f], Christopher G. Chute [b]

SHARP and Beacon Synergies

- SHARP will facilitate the mapping of comparable and consistent data into information and knowledge

SHARPn.org

- 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

SHARPn.org

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
- SHARPn tooling facilitates
  - Normalization/NLP, Phenotyping

SHARPn.org