Fundamental Basis for Data Standards

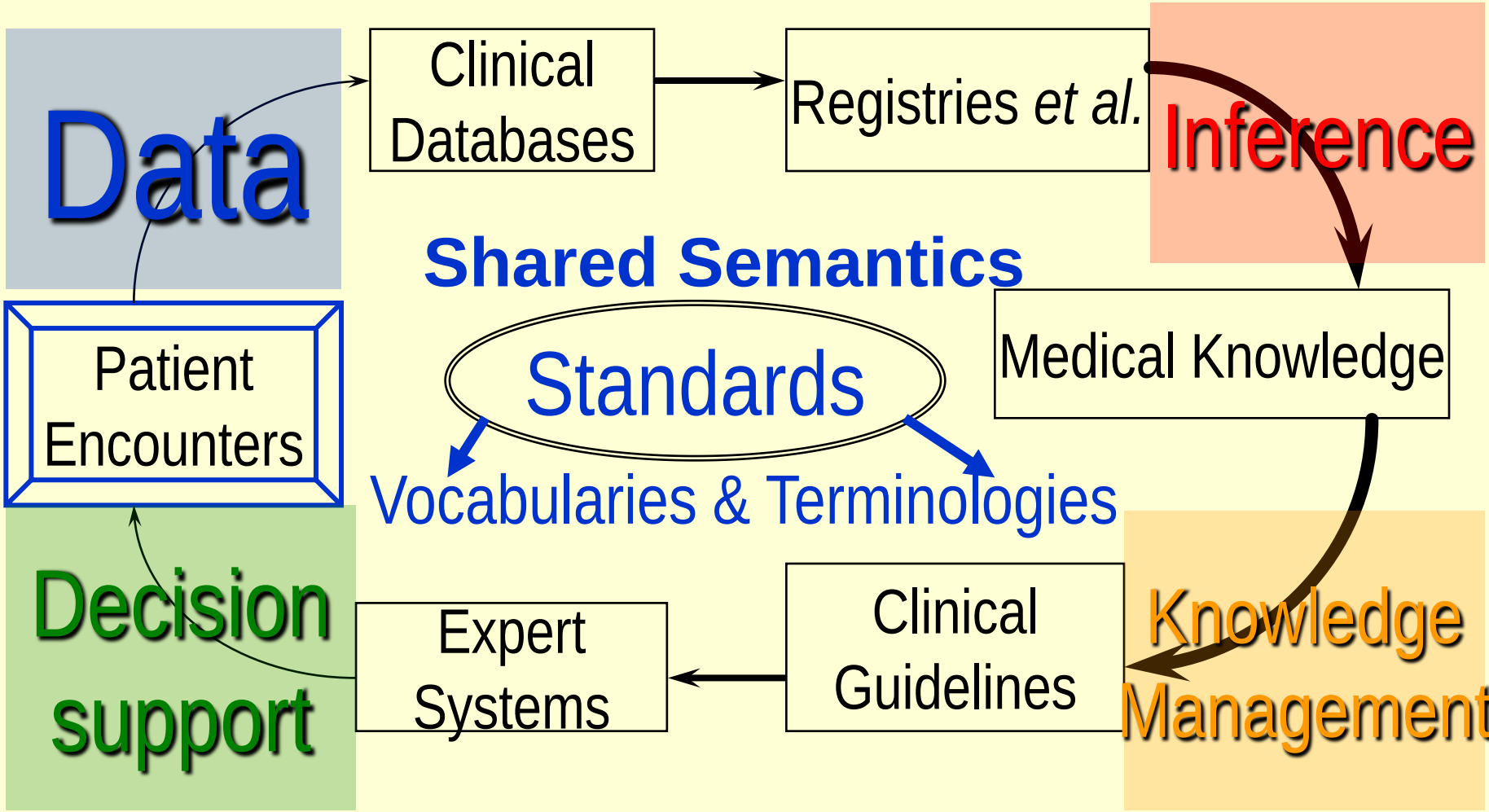
Christopher G Chute MD DrPH, Mayo Clinic
Chair, ICD11 Revision, World Health Organization
Chair, ISO TC215 on Health Informatics

Towards a National Cancer Informatics Program

Bethesda, MD

31 May 2012

From Practice-based Evidence to Evidence-based Practice



Flavors of Data Standards
Sweet-Sour, and the Salt (no bitterness)

- Syntax - Context
  - Data models, small-scale structures
  - Data elements – Lego pieces
- Semantics – Content
  - Vocabulary, terminology, ontology
  - Value sets (bound to models)
- Metadata
  - Provenance
  - Usage and convention

© 2007 Mayo Clinic College of Medicine

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
- Semantics requires context – the model issue

© 2007 Mayo Clinic College of Medicine

Cancer in Context

- There is little unique about “cancer” clinical data
- Cancer-specific clinical standards are:
  - Disruptive, Parochial, Unnecessary
- Cancer Informatics MUST build on clinical informatics community
  - Acknowledge domain-specific extensions

© 2007 Mayo Clinic College of Medicine

Data Granularity Dissonance
Practice vs. Research vs. Public Health

- Clinical data is impressionistic
  - Not protocol driven – quaint rambling text…
  - Detail and consistency is a happy coincidence
  - However, can be holistic; intuit the unanticipated
- Research data is rigid though rigorous
  - Detailed, structured, complete (ideally), coherent
  - May miss the forest for the chlorophyll
- Can there be reconciliation among use-cases?

electronic MEdical Records and GEnomics NHGRI eMERGE (U01) Goals

- GWAS – Genome Wide Association Study
  - 600k Affy chip
- High-throughput Phenotyping
  - Disease algorithm scans across EMRs
  - “catch up” with high throughput genomics
- Generalize Phenotypes across the Consortium
  - Measure reproducibility of algorithms among members
  - Vanderbilt, Northwestern, Marshfield, Group Health Seattle, Mayo

Secondary Use of EHR data Themes & Projects

SHARPn.org

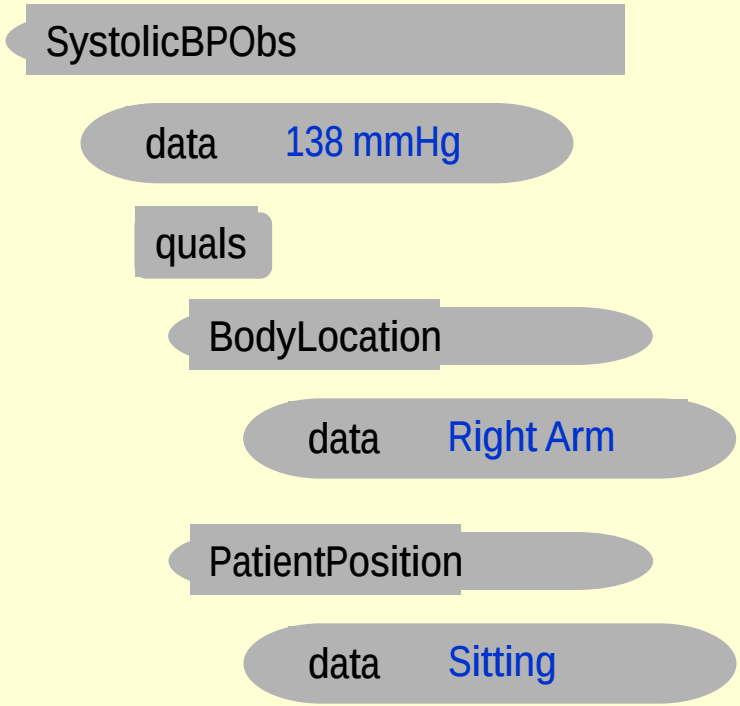
Themes			Projects	Players
Data Normalization	Phenotype Recognition	Data Quality and Evaluation Frameworks	Clinical Data Normalization	IBM, Mayo, Utah, Agilex
			Natural Language Processing (NLP)	Harvard, Group Health, IBM, Utah ,Mayo, MIT, SUNY, i2b2, Pittsburgh, Colorado
			High-Throughput Phenotyping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

Clinical Data Normalization

- Data Normalization
  - Clinical data comes in all different forms even for the same kind of information.
  - 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.

A diagram of a simple clinical model

Clinical Element Model for Systolic Blood Pressure



Clinical Data Element Harmonization

- 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 Clinical Element Models

High-Throughput Phenotyping

- Phenotype - identifying a set of characteristics of about a patient, such as:
  - A diagnosis
  - Demographics
  - A set of lab results
- Phenotyping – overload of terms
  - 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

Where is This Going?

- Standards and interoperability are crucial to the operation and success of any modern scientific enterprise – big science has arrived
- The boundaries among clinical and research standards are eroding
- Information standards, especially vocabularies, are the foundation for scientific synergies
- NCI must collaborate with larger clinical data standards community, especially “Meaningful Use” – parochial standards are not helpful