# More and Better Data for Research:
## *U.S. Health Data Content Standards*

Betsy L. Humphreys
Assistant Director for Health Services Research Information
National Library of Medicine

AcademyHealth
Annual Research Meeting
June 6, 2004

# Presidential Executive Order April 27,2004

With goal of ubiquitous electronic health records and a national health information infrastructure in 10 years:

New National Health Information Technology Coordinator in HHS with responsibility for strategic plan to:

"(i) Advance the development, adoption, and implementation of *health care information technology standards* nationally through collaboration among public and private interests, and *consistent with current efforts to set* health information technology *standards for use by the Federal Government;*

(ii) ...."  http://www.whitehouse.gov/news/releases/2004

# Electronic Health Data Standards
## (including Standard Vocabularies)

- Key element of the health information technology infrastructure for:
  - Effective decision support
  - Safe, evidence-based, and coordinated health care
  - Cost-effective care, assisted by robust market-place and increased/informed choice
  - More efficient clinical, public health, and health services research
  - Timely public health and bioterrorism surveillance

# Types of Health Data

- Administrative health data
    - e.g., health insurance claims

- Clinical data
    - e.g., lab test results, problems, diagnoses, history and physical

- Public health data
    - e.g., disease prevalence, immunization rates, environmental monitoring

*All* potentially relevant to health services research

# Data Content Standards include:

- Data elements, e.g., gender, presenting complaint
- Descriptions of entities, e.g., birth certificate
- Messages, e.g., medication order
- Allowable values for data elements, which can be entire *vocabularies*
- Mappings between different value sets, e.g., between SNOMED and ICD-9-CM
- Information models that define the context in which standards are used
- Survey questions and any coded responses
- Guideline, protocol, and algorithm formats

# Standard Clinical Data

- Enhanced ability to provide access to knowledge where clinical decisions are being made

- Generation of research data as a by-product of health care

- Efficient exchange of data between health care and public health

# Key Acronyms

- **HIPAA** - *Administrative Simplification section of Health Insurance Portability and Accountability Act of 1996*

- **NCVHS** - *National Committee on Vital and Health Statistics, a long-standing (50+ years) advisory committee to HHS, strengthened and expanded by HIPAA*

- **CHI** - *Consolidated Health Informatics project, a cross-agency eGov initiative led by HHS, DOD, and VA*

- **LOINC** - *Logical Observations: Identifiers,Names,Codes*

- **RxNorm** - *Clinical drug nomenclature (ingredient + strength + dose form)*

- **SNOMED CT** - *Systematized Nomenclature of Medicine Clinical Terms - formed by merger of SNOMED and Read*

- **HL7** – *Health Level 7 standard for electronic exchange of clinical data, e.g., lab test results*

# Recommended Steps to Achieving U.S. Health Data Standards (1990-2003)

- ☑ Establish a mechanism for designating U.S. Standards - HIPAA, NCVHS, CHI
- ☑ Pick best available as starting point - NCVHS, CHI
- *Support development, maintenance, and low/no cost distribution
- *Coordinate development of selected standards to achieve non-overlapping, interlocking set
- Broaden participation in standards development
- Promote use and improvement

# Standards Have Been Selected

- U.S. *National* Administrative Standards
  - HIPAA transactions and code sets ncvhs.hhs.gov

- U.S. *Government-wide* Target Clinical Standards
  - CHI message and vocabulary standards www.whitehouse.gov/omb/egov
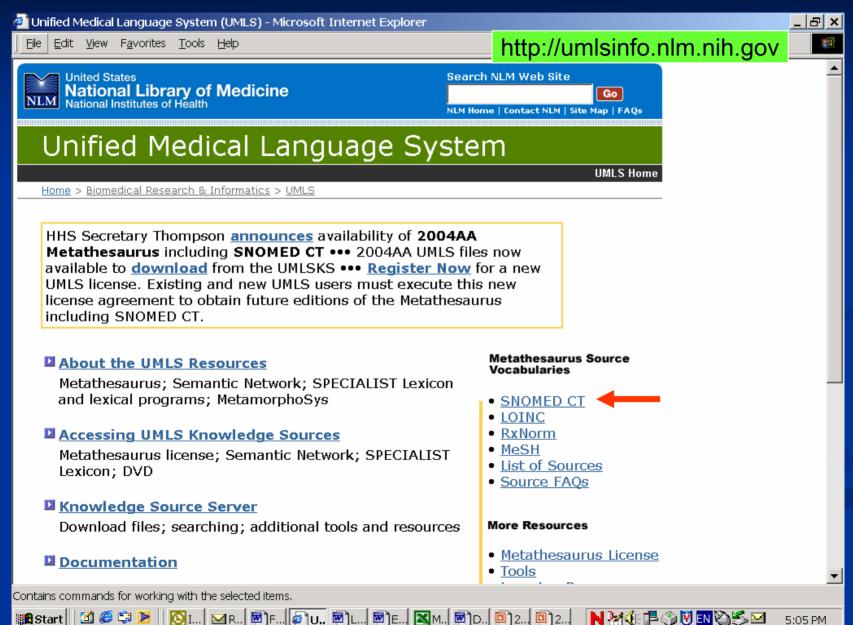
    Choose "Government to Business"

# NLM-led Support for Development and Maintenance

- **1999** – LOINC (lab tests/instrument observations) - contract support

- **2002** – RxNorm (clinical drugs) - direct development

- **2003** – SNOMED CT contract & license for U.S-wide use *(as distributed by NLM in UMLS)*

# NLM No-Cost Distribution
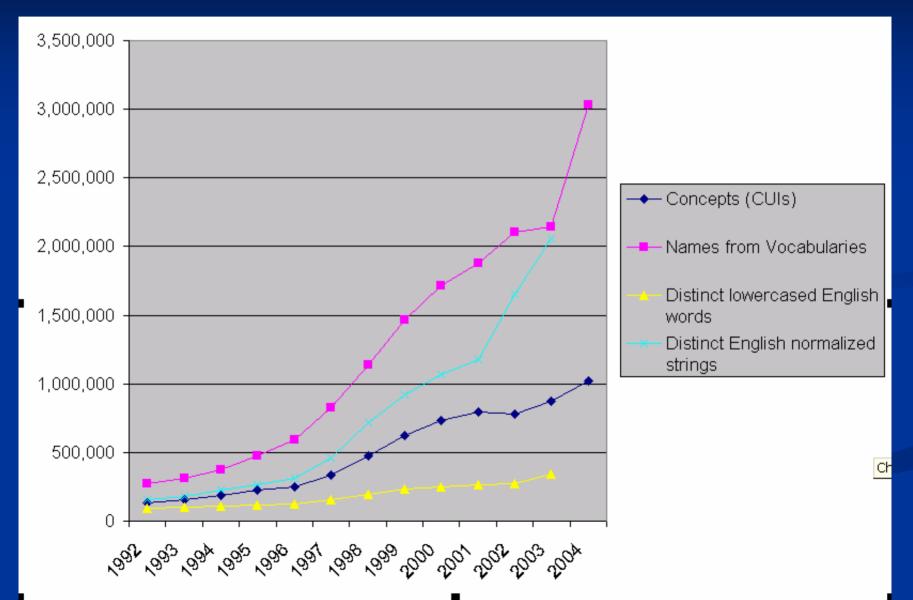
# UMLS® Metathesaurus®
## a *Vocabulary* Database

- **Preserves** the **meanings**, hierarchical connections, and other relationships between terms present in its source vocabularies

- **Adds** certain basic definitional **information** about each of its concepts

- **Establishes new relationships** between concepts and terms from different source vocabularies

- **Distributes many vocabularies in a common, explicit format**

# Other UMLS Resources

- Semantic Network
- SPECIALIST lexicon
- Natural language processing programs
  - In combination with the Metathesaurus, powerful tools for interpretation/indexing of electronic full text

# UMLS Metathesaurus Growth, 1992 -

# 2004AA UMLS Metathesaurus (May 2004)

- ~1,022,000 concepts
- ~2,383,000 unique "strings"
  
  (Eye, Eyes, eye = 3)
- ~3,030,000 source vocabulary terms
- 111 source vocabularies
- 15 different languages

# SNOMED CT® in 2004AA UMLS Metathesaurus

- Active, English content - Jan 31, 2004 edition, including basic mapping to ICD-9-CM
    - Spanish 2004AB;  Inactive 2004AC
- ~300,000 concepts;  ~700,000 strings
- ~37,000 *new* concepts; ~350,000 *new* strings
- UMLS concept view differs: 13% of SNOMED
    - Majority intentional differences
    - Some undiscovered or incorrect synonymy

# Coordinate development and alignment to achieve interlocking set (1)

- UMLS Metathesaurus – common distribution format/mechanism for CHI standards and HIPAA code sets

- 2004AA - UMLS distribution format changes to support
  - Complete "Source Transparency"
  - Easier extraction of subsets for particular purposes
  - Complete "change sets" from previous versions
  - Sophisticated, purpose-specific inter-vocabulary mapping

# Coordinate development and alignment to achieve interlocking set (2)

- Definition of boundaries and relationships
    - SNOMED CT / LOINC / RxNorm
- Alignment of HL7 messages and CHI vocabularies
    - 2004 NLM/HHS contract with HL7 standards organization
- Mappings between vocabularies and classifications
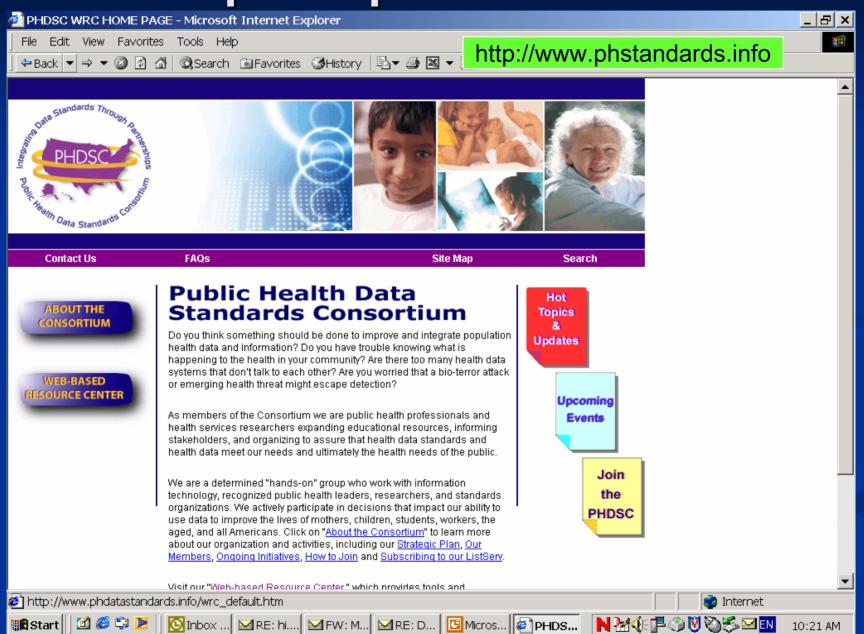
# Mapping Projects planned/underway

- CHI standards – HIPAA code sets
    - SNOMED CT - ICD-9-CM, ICD-10-CM
    - SNOMED CT – CPT
    - LOINC - CPT
- SNOMED CT - "other" vocabularies
    - Medical Dictionary for Regulatory Affairs (MedDRA)
    - International Classification of Primary Care (ICPC)
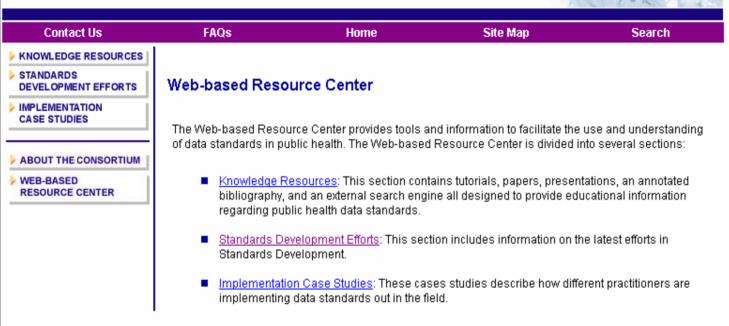    - Medcin
- Will require:
    - Robust testing/validation
    - Alignment of update schedules

# Broaden participation: Public Health

# Web-Based Resource Center

PHDSC

KNOWLEDGE RESOURCES

STANDARDS
DEVELOPMENT EFFORTS

IMPLEMENTATION
CASE STUDIES

ABOUT THE CONSORTIUM

WEB-BASED
RESOURCE CENTER

## Web-based Resource Center

The Web-based Resource Center provides tools and information to facilitate the use and understanding of data standards in public health. The Web-based Resource Center is divided into several sections:

- **Knowledge Resources**: This section contains tutorials, papers, presentations, an annotated bibliography, and an external search engine all designed to provide educational information regarding public health data standards.

- **Standards Development Efforts**: This section includes information on the latest efforts in Standards Development.

- **Implementation Case Studies**: These cases studies describe how different practitioners are implementing data standards out in the field.

Back to Top

**Updated Sunday, 05/02/2004**

# Broaden participation: Clinical Research

# Accelerate Adoption and Use

- Promote testing use of standards by Federal partners, grantees, and contractors
    - *CDC's Public Health Information Network (PHIN) (www.cdc.gov/phin)*
    - *NIH emphasis on use in clinical research networks*

- Encourage manufacturers to include standard identifiers (e.g., LOINC) in device output/test kit packaging

- Collaborate with other HHS agencies to support demonstration/testing

**U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES**
National Institutes of Health

Search ▶

Site Map | Contact

# HIPAA Privacy Rule
Information for Researchers

Home   Dictionary   FAQ   News   Events   Resources

## Educational Materials

**Authorizations**

**Clinical Research**

**HIPAA Privacy Rule Booklet for Research**

**Institutional Review Boards**

**Privacy Boards**

**Research Repositories, Databases**

**Slide Presentations**

## Overview

The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule is the first comprehensive Federal protection for the privacy of personal health information. Research organizations and researchers may or may not be covered by the HIPAA Privacy Rule. This website provides information on the Privacy Rule for the research community.

## HIPAA Resources

- The Privacy Rule - Final Modification (PDF/TXT)

- Office for Civil Rights HIPAA Information (Medical Privacy Home Page)

- Office for Civil Rights HIPAA Guidance (PDF/RTF)

- Office for Civil Rights Summary of the HIPAA Privacy Rule (PDF/RTF)

- Center for Medicare & Medicaid Services HIPAA Information (Covered Entity Decision Tool)

### Highlights

- NEW HRSA Issues a Privacy Rule Resource Guide for HIV Services Providers

- Clinical Research and the HIPAA Privacy Rule

- The Privacy Rule and Public Health

- Research Repositories, Databases

- NIH/AHRQ May 2004 conference on the HIPAA Privacy Rule and Research

Internet

# Take Home Messages

- Health data standards have "arrived"; electronic health records are coming
- Both will affect health care, public health functions, and the data available for HSR
- It's not too early (*or too late*)
  - to contribute to standards development, testing, and refinement
  - to begin to study the impact of standards on health care, public health, and clinical research
- If you work with electronic data, the UMLS resources might be helpful