6

Data Standards

ROBERT E. HOYT

ROBERT W. CRUZ

Learning Objectives

After reading this chapter the reader should be able to:

- Enumerate the reasons data standards are necessary for interoperability
- Understand the importance of clinical summaries such as Continuity of Care Documents (CCDs) or Continuity of Care Records (CCRs)
- Compare and contrast standards used for electronic health records and Meaningful Use

Introduction

According to the Institute of Medicine's 2003 report *Patient Safety: Achieving a New Standard for Care* one of the key components of a national health information infrastructure will be "data standards to make that information understandable to all users."[1]

In order for electronic health records (EHRs), health information organizations (HIOs) and the Nationwide Health Information Network (NHIN) to succeed there needs to be a standard language; otherwise you have a *Tower of Babel*. We use standards every day but often take them for granted. All languages are based on a semantic language standard known as grammar. The plumbing and electrical industries depend on standards that are the same in every state. The railroad industry had to decide many years ago what gauge railroad track they would use to connect railroads throughout the United States.

Interoperability relies on syntax and semantics. Syntax is a concept that is related to the structure of the communication, e.g. HL7 discussed later in the chapter. Semantics is a concept that denotes meaning of the communication e.g. SNOMED also discussed later in the chapter. Data standards can come in many flavors. Standards that focus on communication between multiple systems are referred to as transport standards. The rules that dictate the format of information as it packaged for transport are known as content standards. Individual segments within a content package are governed by a vocabulary. All of these standards are developed after careful study of real world use cases. [2]

Although we have come a long way towards universal standards, we are not there yet. The progress has been slow in part due to the fact that participation in standards determining organizations is voluntary. Data standards have taken on new significance as a result of Meaningful Use objectives and the need for data sharing.

The next sections will discuss the major data standards and how the standards facilitate the transmission and sharing of data. Not all data standards have been included in the following sections and many standards are still a "work in progress."

Common Data Standards

Extensible Markup Language (XML)

- Although XML is not really a data standard it is a data packaging standard that has served as a structural component for domain specific languages for health information exchange. In order for disparate health entities to share messages and retrieve results, a common data packaging standard is necessary
- XML is a set of predefined rules to structure data so it can be universally interpreted and understood
- XML consists of elements and attributes
- Elements are tags that can envelop data and can be organized into a hierarchy. There are no predefined tags
- Attributes help describe the element
- XML messages have headings and message bodies packaging information by wrapping it in layers of "tags." You must write software to send, receive or display these structures
- Below is a simple example where car-lot is the root element and car is a child element. Each car sibling uses attributes to further define the physical model being represented.[3]

```
<car-lot>
 <car make="Ford" model="Mustang">
   <year>1956</year>
   <vin>9216604</vin >
</car>
 <car make="Honda" model="Civic">
   <year>1988</year>
   <vin>9335676</vin>
 </car>
</car-lot>
```

Health Level Seven (HL7)

- A not-for-profit standards development organization (SDO) with chapters in 30 countries
- Health Level Seven's domain is clinical and administrative data transmission and perhaps is the most prolific set of healthcare standards
- "Level Seven" refers to the seventh level of the International Organization for Standardization (ISO) model for Open System Interconnection. This serves to communicate that HL7 messaging

lives in the application layer of the stack, with subordinate layers serving as items in the overall toolkit

- HL7 provides a set of standards for interactions between healthcare data services
- HL7 is a data standard for communication or messaging between:
    - Patient administrative systems (PAS)
    - Electronic practice management systems
    - Lab information systems (interfaces)
    - Dietary
    - Pharmacy (clinical decision support)
    - Billing
    - Electronic health records (EHRs)

Figure 6.1: HL7 example (The vertical bars, called pipes, separate the fields in this example, while the carets separate components)

```
MSH|^~\&EPIC|EPICADT|SMS|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|
EVN|A04|1999122271408|||CHARRIS
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M
NK1||CONROY^MARI^^^^|SPO||(216)731-4359||EC|||||||||||||||||||||||||
PV1||O|168~219~C~PMA^^^^^^^^^^|||277^ALLEN FADZL^BONNIE^^^^|||||||||
```

- The most current version of the HL7 standard is 3.0 but version 2.x is still widely in use by all HIT vendors
- HL7 version 2.x separates messages into processable chunks known as segments which contain fields which contain components
- HL7 version 2.x segments are sewn together into messages of a given type (e.g. Admit Discharge and Transfer | ADT | or Pharmacy Administration [RAS])
- HL7 version 2.x messaging is typically performed over minimal lower layer protocol (MLLP)
- HL7 version 3.0 uses XML for packaging its content
- The Clinical Document Architecture (CDA)
    - A HL7 v3.0 content standard that makes documents human readable and machine processable through the use of XML. CDA is used in EHRs, personal health records, discharge summaries and progress notes. CDA delineates the structure and semantics of clinical documents, consisting of a header and body[6,7]
    - In 2007 HL7 recommended (and HITSP endorsed) the use of the Continuity of Care Document (CCD) standard. The CCD is the marriage of the Continuity of Care Record (CCR) (developed by ASTM International) and the CDA (developed by the HL7 organization)
    - A simpler version of CDA is on its way and is referred to as "greenCDA"

- The info box describes the Health Story Project and templated CDA for including narrative notes into EHRs.

Templated CDA

HEALTH STORY PROJECT

In spite of increasing adoption of EHRs, most patient notes are text and are therefore not discrete data. CDA is a start in the right direction to comply with Meaningful Use.

This HL7 program known as the Health Story Project will match CCD coding patterns and conventions, called "templated CDA." This strategy will help support the transfer of care summaries into an EHR from dictated notes, using CDA templates.[4,5]

- The CCD and CCR:
  - The electronic document exchange standard for the sharing of patient summary information between physicians and within personal health records
  - The CCD has the advantage over CCR of being able to accept free text and being capable of vocabulary specific semantic interoperability. It contains the most common information about patients in a summary XML format that can be shared by most computer applications and web browsers. It can printed (pdf) or shared as html
  - In 2008 CCHIT required EHRs to generate and format CCD documents using the C32 specification for patient demographic information, medication history and allergies.[8-10] The CCD and CCR are both currently listed as interchangeable content standards for achieving Stage 1Meaningful Use
  - The CCD has 17 data content/component modules as part of the C32 standards as noted in Table 6.1. Each module will have additional data elements
  - Dr. John Halamka has posted a sample CCD as an example for others to view[11]

Table 6.1: Data modules of the C32 standard for the CCD

Data Fields					
Patient Demographics	Advance directives	Functional status	Payers	Vital signs	Social history
Purpose	Problems	Alerts	Results	Medical equipment	Procedures
Medications	Encounters	Family history	Immunizations	Plan of care	

Digital Imaging and Communications in Medicine (DICOM)

- DICOM was formed by the National Electrical Manufacturers Association (NEMA) and the American College of Radiology. They first met in 1983 which suggests that early on they recognized the potential benefits of the storage, sharing, and transmission of digital x-rays.
- As more radiological tests became available digitally, by different vendors, there was a need for a common data standard. Similarly, as more EHRs had PACS functionality, DICOM became the standard for images in EHRs.

- While DICOM is a standard, vendors have modified it to suit their proprietary application resulting in lack of true interoperability
- DICOM supports a networked environment using TCP/IP protocol (basic internet protocol).
- DICOM is also applicable to an offline environment.[12]
- "I do Imaging" is a web site that promotes open source DICOM viewers, DICOM converters and PACS clients.[13]

Logical Observations: Identifiers, Names and Codes (LOINC)

- This is a standard for the electronic exchange of lab results back to hospitals, clinics and payers. HL7 is a *content* standard, whereas LOINC is a *vocabulary* standard.
- The LOINC database has more than 30,000 codes used for lab results. This is necessary as multiple labs have multiple unique codes that would otherwise not be interoperable.
- LOINC is divided into lab, clinical and HIPAA portions.
- The lab results portion of LOINC includes chemistry, hematology, serology, microbiology and toxicology.
- The clinical portion of LOINC includes vital signs, EKGs, echocardiograms, gastrointestinal endoscopy, hemodynamic data and others.
- The HIPAA portion is used for insurance claims.
- As an example:
    - The LOINC code for serum sodium is 2951-2; there would be another code for urine sodium.
    - The formal LOINC name for this test is: SODIUM:SCNC:PT:SER/PLAS:QN (component:property:timing:specimen:scale)
- LOINC is accepted widely in the US, to include federal agencies. Large commercial labs such as Quest and LabCorp have already mapped their internal codes to LOINC
- Other standards such as DICOM, SNOMED and MEDCIN have cross references to LOINC
- RELMA is a mapping assistant to assist mapping of local test codes to LOINC codes
- LOINC is maintained by the Regenstrief Institute at the Indiana School of Medicine.[14] LOINC and RELMA are available free of charge to download from www.regenstrief.org/loinc
- For more detail on LOINC we refer you to an article by McDonald.[15]

EHR-Lab Interoperability and Connectivity Standards (ELINCS)

- ELINCS was created in 2005 as a lab interface for ambulatory EHRs and a further "constraint" or refinement of HL7 standards.
- Traditionally, lab results are mailed or faxed to a clinician's office and manually inputted into an EHR. ELINCS would permit standardized messaging between a laboratory and a clinician's ambulatory EHR.
- Standard includes:
    - Standardized format and content for messages

- Standardized model for transport of messages
- Standardized vocabulary (LOINC)

- The Certification Commission for Healthcare Information Technology (CCHIT) has proposed that ELINCS be part of EHR certification.
- HL7 plans to adopt and maintain the ELINCS standard.
- California Healthcare Foundation sponsored this data standard.[16]

IEEE 11073

- Data standards are needed for information to be sent from a medical device to an EHR or hospital information system
- This is a fundamental standard for medical device connectivity and data exchange but is not widely used
- HL7 version 2.x is used for data transfer but only supplies the syntax and not the semantics
- Other initiatives are being developed to solve this interoperability problem
  - Integrating the Healthcare Enterprise-Patient Care Device (IHE-PCD) Workgroup has developed use case profiles to support integration, alerts and implantable devices
  - Medical Device Plug and Play Interoperability Program's Integrated Clinical Environment will develop a solution like IHE-PCD that will be based on IEEE 11073
  - IEC 80001 is standard under development to address devices in a networked environment
  - Continua Health Alliance focuses on home healthcare devices[17]

RxNorm

- RxNorm is the recommended standard for medication vocabulary for clinical drugs and drug delivery devices, developed by the National Library of Medicine (NLM).
- Each commercial drug vocabulary company e.g. First Data Bank provides medication concept identifiers to the NLM which are then mapped to the concepts in the RxNorm vocabulary.
- Supports interoperability among organizations that deal with clinical drugs.
- The standard includes three drug elements: the active ingredient, the strength and the dose.
- RxNorm is the standard for e-prescribing and will support Meaningful Use.
- RxNorm encapsulates other drug coding systems, such as National Drug Code (NDC).
- The standard only covers US drugs at this point.
- An example of RxNorm: 311642 Methylcellulose 10 MG/ML Ophthalmic Solution.[18]

National Council for Prescription Drug Programs (NCPDP)

- NCPDP is a pharmacy related SDO for exchange of prescription related information
- Script (v10.10) is for communication between physician and pharmacist

- Other standards: batch standard, billing standard, formulary and benefit standard, prescription file transfer standard and universal claim form standard [19]

Accredited Standards Committee (ASC) X12

- A standard for electronic data interchange (EDI) or the computer-computer exchange of business data
- Standard is used in healthcare, transportation, insurance and finance industries [20]

Systematized Nomenclature of Medicine: Clinical Terminology (SNOMED-CT)

- SNOMED is the clinical terminology or medical vocabulary commonly used in software applications, including EHRs
- SNOMED covers diseases, findings, procedures, drugs, etc.; a more convenient way to index and retrieve medical information
- The vocabulary provides more clinical detail than ICD-9 and felt to be more appropriate for EHRs
- SNOMED is also known as the International Health Terminology
- This standard was developed by the American College of Pathologists. In 2007 ownership was transferred to the International Health Terminology Standards Development Organization www.ihtsdo.org
- SNOMED will be used by the FDA and the Department of Health and Human Services
- This standard currently includes about 1,000,000 clinical descriptions
- Terms are divided into 19 hierarchical categories
- The standard provides more detail by being able to state condition A is due to condition B
- SNOMED concepts have descriptions and concept IDs (number codes). Example: open fracture of radius (concept ID 20354001 and description ID 34227016)
- SNOMED CT also defines two types of relationships:
    - "Is a" connects concepts within the same hierarchy. Example: asthma "is a" lung disease
    - "Attribute" connects concepts in different hierarchies. Example: asthma is associated with inflammation
- SNOMED links (maps) to LOINC and ICD-9/10
- SNOMED is currently used in over 40 countries
- EHR vendors like Cerner and Epic are incorporating this standard into their products
- There is some confusion concerning the standards SNOMED and ICD-9; the latter used primarily for billing and the former for communication of clinical conditions [21-23]
- A study at the Mayo Clinic showed that SNOMED-CT was able to accurately describe 92% of the most common patient problems [24]

- SNOMED-CT Example: Tuberculosis

  DE-14800

  . . . .

  . . . .

  . . . Tuberculosis

  . . Bacterial infections

  . E = Infectious or parasitic diseases

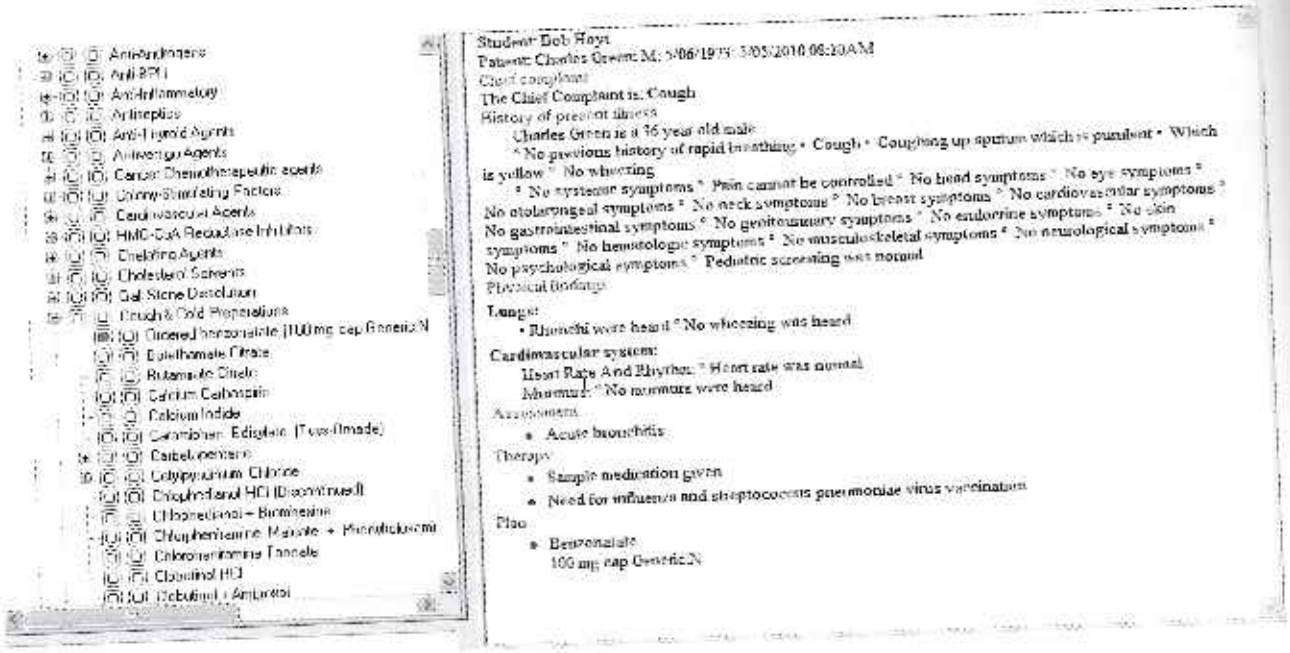
  D = disease or diagnosis

MEDCIN®

MEDCIN ® was developed by Medicomp in the 1980s as a proprietary medical vocabulary. In 1997 it was released as a national standard. MEDCIN® cross-references to many of the other standards already discussed. The nomenclature consists of about 270,000 clinical concepts organized into categories: symptoms, history, physical exam, tests, diagnosis and therapy. Each finding is associated with a numerical code, up to seven digits, so the results are structured or codified. Unlike SNOMED, MEDCIN® findings can link to symptoms, exam, therapy and testing. The knowledge base also includes 600,000 synonyms, allowing look-ups under different terms. MEDCIN® is used by several EHR systems, to include the DOD's AHLTA.

The disadvantages of this system are the fact that there is a substantial learning curve to be able to search for all of the necessary MEDCIN® terms in order to create a completely structured note. Second, the note that is created tends to be poorly fluent and not like dictation (Figure 6.2). For that reason, Medicomp developed CliniTalk™ which is a voice to text option that means that a clinician can dictate and the end result is structured data. [25]

Figure 6.2: Simple note created with MEDCIN® (Courtesy MEDCIN®)

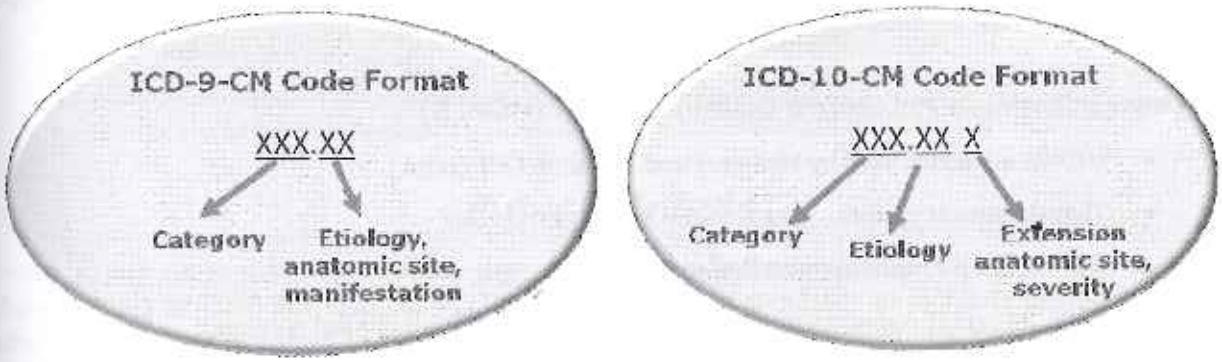


International Classification of Diseases 9th revision and ICD-10

ICD-9 is published by the World Health Organization to allow mortality and morbidity data from different countries to be compared. The basic ICD-9 code, for example diabetes, is 250.00 (three digits). If it covers more detail such as diabetes with kidney complications, it becomes 250.40 (four or five digits). An online web site lists common ICD-9 CM codes. [26]Although it is the standard used to provide a diagnosis for billing over the past 30 years, it is not ideal for distinct clinical diseases.

ICD-10 was endorsed in 1994 but not used in the US. The Federal government set October 2013 as the launch date for ICD-10.[27] ICD-10 will provide a more detailed description with seven rather than five digit codes. ICD-10 CM (clinical modifications) was developed by the CDC for use in all healthcare settings. ICD-10 PCS (procedure coding system) was developed by CMS for inpatient settings only and will replace ICD-9 CM, volume 3. [28] ICD-10 would result in about 141,000 codes instead of the 17,000 codes currently used. A study by Blue Cross/Blue Shield estimated that adoption of ICD-10 would cost the US healthcare industry about $14 billion over the next two to three years. The more digits included generally results in higher reimbursement. [29, 30] Figure 6.3 demonstrates the differences between ICD-9 and ICD-10 formats.

Figure 6.3: ICD-9 and ICD-10 Code Formats



Current Procedural Terminology (CPT)

- CPT is used for billing the level and complexity of service rendered
- Standard was developed, owned and operated by the American Medical Association (AMA) for a fee
- A CPT code is a five digit numeric code that is used to describe medical, surgical, radiology, laboratory, anesthesiology, and evaluation/management services of physicians and hospitals. There are approximately 7,800 CPT codes ranging from 00100 through 99499. Two digit modifiers may be appended to clarify or modify the description of the procedure. The most recent CPT book is version 4
- CPT Codes are published in two versions: the first is for physicians and the second is for hospitals.
- Example CPT code: office visit, established patient, moderate complexity = 99214 [31]

Unified Medical Language System (UMLS)

- Developed by the National Library of Medicine (NLM) to facilitate semantic interoperability among computer systems. Consists of files and software

- UMLS could be used to develop applications, such as electronic health records, dictionaries, language translators, etc.
- UMLS is unique, in that is can link health information, medical terms, drug names, and billing codes across different computer systems
- Access to the UMLS is free but requires registration
- UMLS concept mappings are *as-is* and must be editorialized in order to safely embed in a functional system
- The UMLS Knowledge Sources consist of three components
    - Metathesaurus is a very large multi-lingual vocabulary database with information and relationships (concepts with meaning) related to healthcare: CPT®, ICD-10-CM, LOINC®, MeSH®, RxNorm, and SNOMED CT®
    - Semantic Network categorizes all concepts within the Metathesaurus e.g. anatomical structures, biological function, physical objects, etc.
    - SPECIALIST Lexicon is an English lexicon that includes biomedical terms that support natural language progressing (NLP) [32]

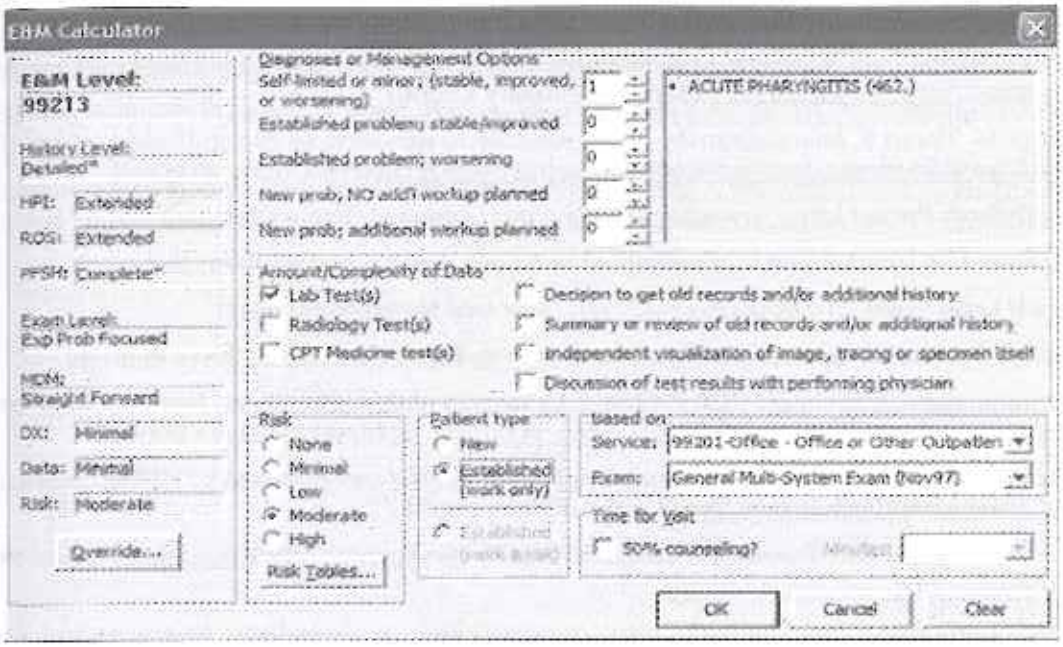
Healthcare Common Procedure Coding System (HCPCS)

- HCPCS are codes used by Medicare and based on CPT codes
- They document medical, surgical and diagnostic services
- HCPCS Level I codes are identical to CPT codes
- HCPCS Level II codes are used by medical suppliers and not clinical services
- More information is available from CMS [33]

Evaluation & Management (E&M) Codes

- In order to bill for a patient visit, ICD-9 CM and CPT codes are selected to best represent the visit. It is up to the clinician to provide documentation to prove the level of the visit
- As an example, if a clinician chooses to select CPT code 99204 for a new outpatient patient visit, they must document that the problems are of moderate to high severity, the physicians spends about 45 minutes face-to-face and the E&M requires these key components: comprehensive history and physical exam and medical decision making of moderate complexity. This implies that an excellent history and physical exam are documented and the problems discussed were moderately complex
- Many EHRs have E&M calculators to help assist the clinician in determining the level of service. This is made easier if templates are used because clicking on history and physical exam elements can calculate an E&M code in the background [34]
- Figure 6.4 shows a typical E&M calculator that is part of an EHR. Note: this is an established patient, the E&M level is in the upper left, the diagnosis and ICD-9 code (462) are in the upper right. Multiple fields are available to input the complexity of the visit so the E&M code can be manually or automatically calculated

Figure 6.4: E&M calculator as part of Healthmatics EHR (Courtesy www.network-systems.com**)**



Future Trends

We can expect more data standards as time goes by and further refinement of all existing standards. Subcommittees of ONC are working hard to harmonize data standards to facilitate health information exchange. Decisions will have to be made about what standards will be mandatory for electronic health records. For example, we anticipate that SNOMED will be the primary medical vocabulary of choice for electronic health records in the near future. Meaningful Use stages 2 and 3 may direct all new EHR data standards and vocabularies.

Key Points

- Data standards play a major role in accomplishing interoperability
- We are slowly moving towards industry wide standards, such as the Continuity of Care Document
- Meaningful Use is a strong driver of data standards development

Conclusion

Data standards are critical for interoperability between disparate technologies and organizations. Without agreed upon standards for content and terminology, true semantic interoperability is next to impossible. Multiple standards developing organizations have proposed standards that are being tested, harmonized and updated for application in the field of medicine. Standards are important to exchange clinical data, as well as, administrative and financial data. Standards are essential for exchange of information between electronic health records, health information organizations and the Nationwide Health Information Network. Data standards are on the radar screen as a result of need to meet Meaningful Use and work by groups such as the Health Information Technology Standards Committee.