

Collect Once, Use Many: Enabling the Reuse of Clinical Data through Controlled Terminologies

Save to myBoK

by James J. Cimino, MD, FACMI, FACP

Capturing patient information with controlled terminologies can enable its reuse for a wide variety of purposes, from bed assignments to reporting to clinical alerts.

The requirements for collecting information about patients and the process of their care have grown constantly since Florence Nightingale bemoaned the lack of adequate health records, through Larry Weed's promotion of the problem-oriented medical record, and into the age of documentation for prospective payment, utilization review, and healthcare report cards.

Despite the corresponding increase in the use of computers to capture, store, and retrieve the information to meet these requirements, a great deal of redundancy exists in the process of the actual collection, while the goal of reusing the data for multiple purposes remains elusive.

For example, clinicians dutifully record histories, physicals, impressions, and plans in their patients' medical records, but they must then take time to fill out additional forms to report medication lists, problem lists, and reasons for visits. Healthcare institutions employ entire departments devoted to abstracting records to provide documentation for reimbursement.

Computers can best help us with reuse of data when data are represented in meaningful ways. The mere appearance of the word "pneumonia" in a health record does not tell us if the patient carries the diagnosis or has had the diagnosis excluded, and the absence of the word does not guarantee the absence of the diagnosis. Instead of relying on the textual portion of the record, computer systems make use of data coded with controlled terminologies, the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) being the most widely used example. This article illustrates how capturing patient data with such controlled terminologies can support the reuse of patient data for a wide variety of purposes.

Sample Case Record

The patient is a 50-year-old woman who presents to the emergency room with the chief complaints of fever, cough, and chest pain.

The patient has a history of hypertension and diabetes mellitus, with a 70 pack-year smoking history, and was in her usual state of good health until four days prior to admission, when she noted the onset of fever, chills, and cough productive of green sputum. These symptoms continued until the day of admission, when the patient noted the onset of dull, retrosternal chest pain, made worse with exertion, so she came to the emergency room. The patient reports that she takes a pill for blood pressure and a pill for diabetes but does not recall the names. She states that she once got a rash after taking Bufferin.

On arrival to the emergency room, physical examination revealed a well-developed, well-nourished woman in moderate respiratory distress. Vital signs showed a pulse of 80 and regular, a blood pressure of 160/100, a respiratory rate of 22 breaths per minute, and an oral temperature of 100.8°F. Examination of her chest revealed rales and rhonchi on the right side. The remainder of the examination was normal.

Laboratory examination included a serum sodium of 125 mEq/L (normal [136–146]); serum glucose of 200 mg/dL (normal [70–105]), a positive serum troponin, and a negative urine glucose. An electrocardiogram showed ST segment depression in the anterior leads. A chest x-ray showed an infiltrate in the right lower lobe. The patient was admitted to the hospital.

A sample case record showing a variety of typical clinical data.

Examples of the information's reuse appear in the text.

Case Study

Consider the abbreviated patient record above. This short record contains a wealth of data that can be used to understand the patient's condition and guide her care. Humans reading this case would have no trouble deciding which room the patient should be admitted to or picking out her blood sugar results. With a little additional medical knowledge, humans can also suggest that the patient should be given prophylactic aspirin therapy because of evidence of an acute myocardial infarction.¹ This would be advised with caution, owing to a possible aspirin allergy. A computer, on the other hand, would be hard pressed to find the specific words it might expect for such tasks in the record.

An Example: Bed Allocation

One of the simplest examples of data reuse is extracting the patient's age and gender from her record and providing it to the hospital's admission/discharge/transfer (ADT) system to automate the process of assigning the patient to a hospital bed. However, if the system has a rule in the form:

If PATIENT.GENDER = "FEMALE" Then Assign Patient to EMPTY.FEMALE.BED

it will fail if the patient's gender is recorded as (in our case) "woman."

Instead of relying on the ability of the ADT system to pick the gender information from the text, we can record our data using a controlled terminology. In this case, we can imagine a very simple terminology consisting of two terms, male and female, each of which is represented by a unique identifier. (This example ignores situations of ambiguous gender, transgender, and other deviations from the male-female dichotomy.)

Instead of recording the patient's gender in the text, we would record it using a choice from the terminology, perhaps using a pick list or other data-entry mechanism. The ADT system can then use its logic to match the appropriate unique identifier unambiguously with an appropriate bed assignment.

Solutions Using Controlled Terminology

Even the simple example of reuse of gender information illustrates some of the challenges in using controlled terminologies: coverage of the domain and capturing data in a coded form.

We can consider other, more complex examples for which the payoff is even greater than automated bed assignment.

Laboratory Data Summarization

The primary use of laboratory data is to report results to clinicians and patients; the most common reuse is the summarization of those data to show patterns and trends. If we wished to summarize our patient's blood sugar results, we would want to extract "serum glucose" and "plasma glucose" from the record but exclude "urine glucose."

The good news is that laboratory systems generally use controlled terminologies to report their data. The bad news is that the appropriate summarization of results often requires including multiple term codes in each category, so the summarization programs need to know which codes to aggregate. Furthermore, laboratories often make changes to their terminologies, requiring the summarization programs to stay current with these changes.

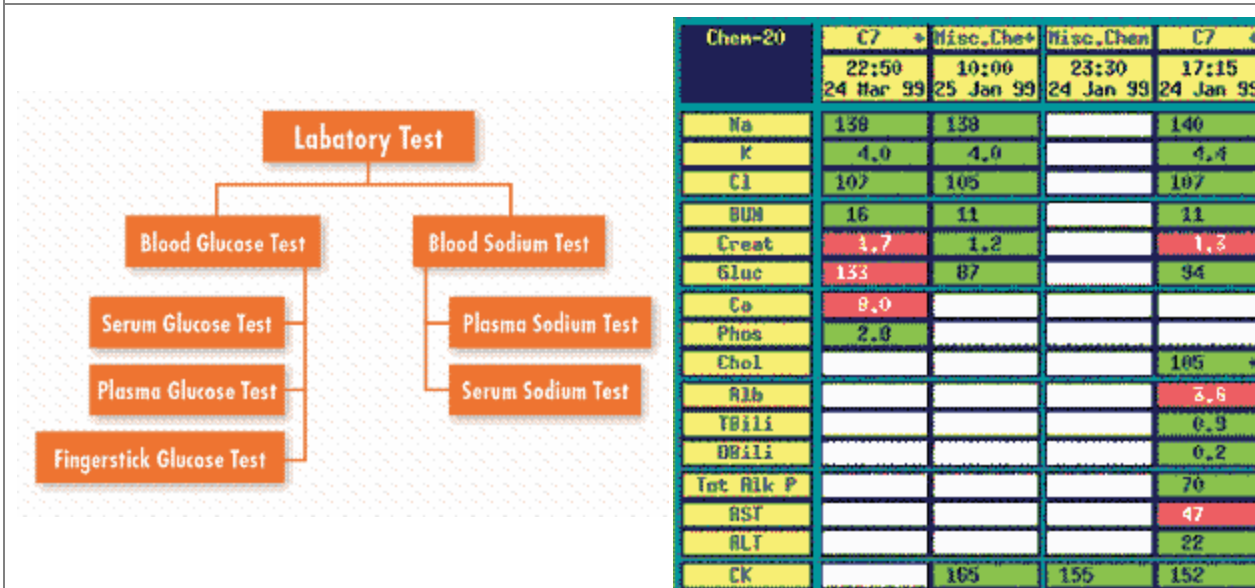
Controlled terminologies can help with these tasks when they include classification systems that identify how individual coded terms can be grouped. The [figure](#) [below] depicts a sample of terms from a laboratory terminology and shows how classification can support summarization of patient data by creating hierarchical relationships of the test terms. Controlled terms for new tests can be added to the hierarchy, allowing results to be aggregated automatically.

Clinical Guidelines

Computerized clinical guidelines are often cited as the tools that will help deliver evidence-based patient care, and electronic medical records are held out as the means to enable the delivery of appropriate guidelines to the point of care.² However, guidelines depend on the availability of coded patient data to satisfy their criteria.

Like the ADT system, some guidelines require age and gender information to make healthcare maintenance recommendations such as mammograms and Pap smears. A more complex recommendation, such as the suggestion that the patient in our example should be given aspirin, would require coded terms for “exertional chest pain,” “ST segment depression,” and “positive serum troponin” to determine that the patient meets criteria for possible myocardial infarction. Thus, terminologies must cover not only all the terms in specific domains, such as gender or laboratory tests, but represent multiple domains across the spectrum of patient care data.

Grouping Individual Terms for Summarizing Patient Data



A controlled terminology (left) showing hierarchical relationships among laboratory test terms. The hierarchy can be used to aggregate patient data for the summary displays on the right. For example, results for plasma glucose tests, serum glucose tests, and fingerstick glucose tests can be organized into a single row of a display, as shown on the right. If the laboratory decides to add a new test (e.g., whole blood glucose), the controlled term for the test can be added to the terminology as a child of blood glucose test, and the test’s results will automatically be aggregated into the appropriate row of the display.

Clinical System Alerts

Clinical information systems are effective for going beyond simple recommendations to monitor patient data in real time and trigger alerts to dangerous situations.³ Whether the alert involves worrisome trends in patient data or errors in clinicians’ orders, the data must be coded using controlled terminologies to be effective.

If, in our sample case, the responsible physician or nurse practitioner decides that the patient should be given aspirin (perhaps because of a computerized guideline), we would hope that the computerized order entry system would issue a warning about the patient’s possible allergy. Although a rule such as:

If ORDER=“aspirin” and ALLERGY=“Bufferin” Then Print “Warning!”

might work, trying to include a rule for every order and allergy permutation would be impractical. A more generic rule such as:

If ORDER=ALLERGY Then Print “Warning!”

would fail, since the order for “aspirin” would not match the word “Bufferin” that appears in the text.

However, if we represent the patient's allergies and ordered medications by using a controlled terminology, we can keep track of additional knowledge associated with the terms that can be used by computerized reasoning systems. In this example, we might include drug ingredient information in our terminology, as in the figure on the following page, to support the inference that aspirin should not be ordered for patients with Bufferin allergy. The resulting rule might then be:

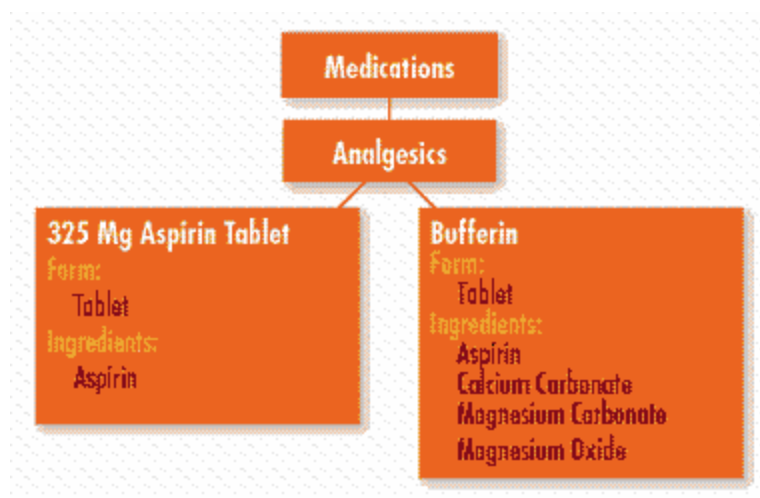
If {any ingredient of ordered drug} = {any ingredient of allergic drug} = Then Print "Warning!"

Automated Retrieval of Health Information

Besides satisfying information needs, the data in a patient's record can trigger new information needs, such as the need for explanations of the patient findings or appropriate therapies for patient conditions. Although a variety of information resources are available to help resolve such needs, the automated reuse of patient data to help obtain answers from these resources is complicated by the fact that the resources typically do not speak the same language as the patient's record.

In our example, the patient has a serum sodium level of 125 mEq/L and a serum glucose level of 200 mg/dL. Literature databases such as PubMed and diagnostic expert systems such as DXplain can provide information about the causes of such findings.^{4,5} However, they index their information with controlled terms that differ from those in the patient's record. Controlled terminologies from different systems can be integrated with each other, as they are in the National Library of Medicine's Unified Medical Language System.⁶ Automated translation then can be brought to bear to support reuse of patient data for automated knowledge retrieval. The figure at right shows an example of how automated terminology translation is used by automated links, called infobuttons, to carry out such retrievals.⁷

Increasing Detail for Clinical Alerts



A controlled terminology for medications that includes drug ingredient information. A terminology such as this could be used for comparing medication orders to patient allergies by using a rule such as:

If {any ingredient of ordered drug} = {any ingredient of allergic drug} Then Print "Warning!"

A Look Ahead

The examples of data reuse described here are a small sample of the ways in which computer systems can use available information to improve the healthcare process without redundant data collection. In each case, the key to success is the representation of the information using controlled terms. Without such representation, the data are merely strings of letters and numbers to be stored and displayed; with such representation, the terms used to code the data convey semantic information that can support reasoning about the data and the patient to whom they relate.

There are significant challenges to representing patient data with controlled terminologies, including lack of availability of high-quality terminologies, resistance to the adoption of standards for representation, and inadequate mechanisms for capturing the

data in coded form. Although ICD-9-CM is a widely adopted standard, it suffers from a number of technical limitations that render it unfit for capturing data suitable for reuse, including its inability to allow multiple classifications of terms and its lack of semantic knowledge beyond term names and codes.⁸

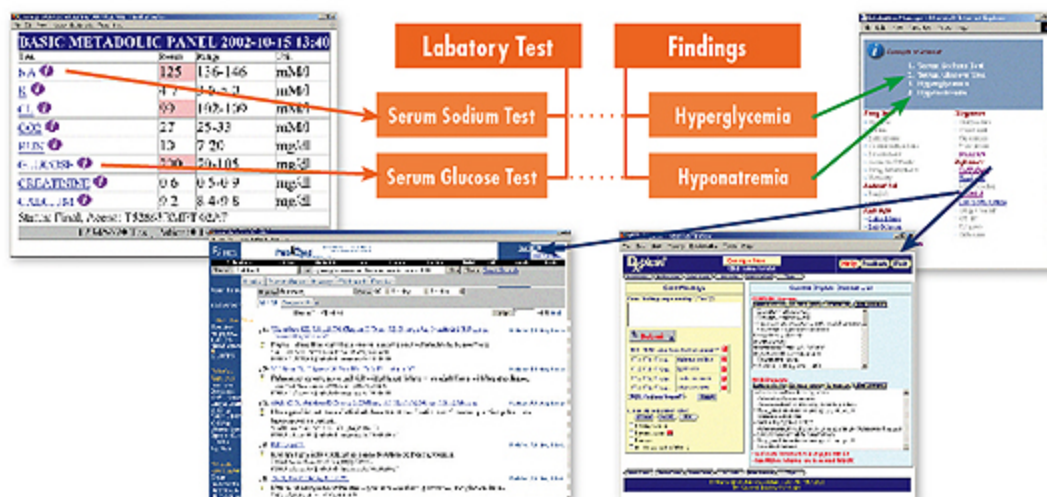
Fortunately, several standards efforts have begun to produce terminologies that can support data reuse, including the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) for clinical findings, Logical Observation Identifiers Names and Codes (LOINC) for laboratory tests, and RxNorm for medications.^{9,10,11} Terminologies such as these address the issues of domain coverage, multiple hierarchies, and semantic representation that support tasks such as the reuse examples described above. Besides their intrinsic properties, each of these terminologies has been included in the Unified Medical Language System Metathesaurus, which can support translation of terms across multiple terminologies and domains.¹²

There is significant resistance to the adoption of these terminologies as standards; controversy and opposition currently impede the adoption of even ICD-10-CM over ICD-9-CM.¹³ However, HIPAA mandates the use of standards for the transmission of health information, and the National Committee on Vital and Health Statistics is in the process of making standards recommendations to the secretary of Health and Human Services, indicating that standards adoption is inevitable.^{14,15}

The ability to capture patient information with controlled terms remains difficult, but user interface design and natural language processing are reducing some of the barriers to system adoption.^{16,17} The currently increasing use of clinician order entry systems is one way in which important clinical data (diagnostic tests and medications) are being captured in coded, if not yet standard, form.

The ability to reuse patient data to improve healthcare processes need not await complete encoding of the patient record. A great deal can be done, and is being done, with the data we have coded thus far. As we code additional parts of the record, we will have greater capability to build systems that reuse data.¹⁸ As coding systems become more standardized, the solutions that address these uses can be standardized as well.

Integrating Terminologies to Link Patient Data to Literature Databases



Automated retrieval of health information from a patient's electronic medical record. Infobuttons in a clinical information system (top left) use controlled terms for the laboratory tests (red arrows) that are related to finding terms in the controlled terminology (top center). The finding terms can be passed (green arrows) to an infobutton manager (top right), which uses the finding terms to retrieve information (blue arrows) from a literature database (PubMed, bottom left) and a diagnostic expert system (DXplain, bottom right).

Notes

1. Anand, S.S., and S. Yusuf. "Oral Anticoagulants in Patients with Coronary Artery Disease." *Journal of the American College of Cardiology* 19, no. 41 (4 Suppl S) (2003): 62S–69S.

2. McDonald, C.J. "The Barriers to Electronic Medical Record Systems and How to Overcome Them." *Journal of the American Medical Informatics Association* 4, no. 3 (1997): 213–21.
3. Kuperman, G.J., J.M. Teich, T.K. Gandhi, and D.W. Bates. "Patient Safety and Computerized Medication Ordering at Brigham and Women's Hospital." *Joint Commission Journal on Quality Improvement* 27, no. 10 (2001): 509–21.
4. McEntyre, J., and D. Lipman. "PubMed: Bridging the Information Gap." *Canadian Medical Association Journal* 164, no. 9 (2001): 1317–19.
5. Barnett, G.O., J.J. Cimino, J.A. Hupp, and E.P. Hoffer. "DXplain: An Evolving Diagnostic Decision-support System." *Journal of the American Medical Association* 258, no. 1 (1987): 67–74.
6. Lindberg, C. "The Unified Medical Language System (UMLS) of the National Library of Medicine." *Journal of the American Medical Record Association* 61, no. 5 (1990): 40–42.
7. Cimino, J.J., G. Elhanan, and Q. Zeng. "Supporting Infobuttons with Terminological Knowledge." *Proceedings: A Conference of the American Medical Informatics Association Annual Fall Symposium* (1997): 528–32.
8. Cimino, J.J. "Desiderata for Controlled Medical Vocabularies in the Twenty-first Century." *Methods of Information in Medicine* 37, no. 4–5 (1998): 394–403.
9. Spackman, K.A. "SNOMED CT Milestones: Endorsements Are Added to Already-impressive Standards Credentials." *Healthcare Informatics* 21, no. 9 (2004): 54–56.
10. Huff, S.M., et al. "Development of the Logical Observation Identifier Names and Codes (LOINC) Vocabulary." *Journal of the American Medical Informatics Association* 5, no. 3 (1998): 276–92.
11. Cimino, J.J., and X. Zhu. "The Practical Impact of Ontologies on Biomedical Informatics." *Methods of Information in Medicine* 45 Suppl 1 (2006): 124–35.
12. Humphreys, B.L., D.A. Lindberg, H.M. Schoolman, and G.O. Barnett. "The Unified Medical Language System: An Informatics Research Collaboration." *Journal of the American Medical Informatics Association* 5, no. 1 (1998): 1–11.
13. National Committee on Vital and Health Statistics (NCVHS). Subcommittee on Health Data Needs, Standards and Security. "Day 2, April 16, 1997." Transcript. Available online at <http://ncvhs.hhs.gov/970416tr.htm>.
14. NCVHS. "Uniform Data Standards for Patient Medical Record Information." 2000. Available online at www.hipaadvisory.com/regs/ncvhsexecsum.htm.
15. NCVHS. "Report to the Secretary of the US Department of Health and Human Services on Uniform Data Standards for Patient Medical Record Information." July 6, 2000. Available online at www.ncvhs.hhs.gov/hipaa000706.pdf.
16. Rosenbloom, S.T., et al. "Interface Terminologies: Facilitating Direct Entry of Clinical Data into Electronic Health Record Systems." *Journal of the American Medical Informatics Association* 13, no. 3 (2006): 277–88.
17. Tange, H.J., et al. "Medical Narratives in Electronic Medical Records." *International Journal of Medical Informatics* 46, no. 1 (1997): 7–29.
18. Cimino, J.J. "Data Storage and Knowledge Representation for Clinical Workstations." *International Journal of Bio-Medical Computing* 34, no. 1–4 (1994): 185–94.

James J. Cimino (ijc7@columbia.edu) is a professor in the Departments of Biomedical Informatics and Medicine at Columbia University College of Physicians and Surgeons, New York.

Article citation:

Cimino, James J.. "Collect Once, Use Many: Enabling the Reuse of Clinical Data through Controlled Terminologies" *Journal of AHIMA* 78, no.2 (February 2007): 24–29.