The Need for a Concept-Based Medication Vocabulary as an Enabling Infrastructure in Health Informatics

W. David Sperzel, MD, MS; Carol A. Broverman, PhD; Joan E. Kapusnik-Uner, PharmD; Joseph M. Schlesinger

Health Informatics Group, First DataBank, Inc., San Bruno, CA

ABSTRACT

Users of drug information typically focus their attention at different levels of description in different situations, such as medication ordering or dispensing. Computer systems utilizing drug information to support such activities must accommodate these multiple perspectives. This paper presents an approach to conceptualizing drug descriptions at multiple levels and outlines key features of an underlying information model that can serve as the basis for a concept-oriented medication vocabulary. These features include dose forms, routes of administration, as well as links to multiple drug classification schemes and medical problems. Implementation, standards, and maintenance issues related to the model are also discussed.

MULTI-LEVEL DRUG DESCRIPTIONS

The late M. S. Blois argued that medical information must be considered at multiple hierarchical levels, ranging from descriptions of atoms and molecules to descriptions of the patient as a whole[1]. The idea of multiple levels of description can be also be applied quite productively to the world of drugs. The question "What is a drug?" can be answered at different descriptive levels, depending on the context in which the question is asked. A pharmacologist interested in mechanism of action may think of a drug at the molecular level, where it interacts with a receptor site. A prescribing physician may be primarily concerned with selecting a drug by its generic or brand name and specifying an appropriate dosing regimen for a particular patient, while a dispensing pharmacist may focus on filling the prescription with an appropriate packaged product from the pharmacy's inventory. Various kinds of clinical and financial data must be linked to descriptive drug information at the appropriate level. For example, clinical information about drug interactions should probably be linked at a level that merely identifies the sets of ingredients in various drug preparations, while cost information only has meaning for packaged drug products available for purchase.

In identifying different levels of description that can be usefully applied to drugs, we are actually creating a series of abstractions to help us focus on those drug characteristics that are relevant to particular points of view. Clinical decision-makers typically consider drugs at more abstract levels of description than people who are, for example, dealing with inventory or pricing issues.

The left side of Figure 1 illustrates an approach to conceptualizing drug descriptions at multiple levels. The notation, adapted from Blois[2], consists of a nominal separated from a series of attributes by two vertical bars ("||"). An ellipsis ("...") is used to indicate that the list of attributes is not necessarily complete. The abstraction process is reflected by the decreasing number of attributes at each level as we descend the hierarchy. Each level depicted in the figure has a many-to-one relationship with the levels below it. The *packaged drug product*, *manufactured formulation*, and *ingredient* levels of description correspond directly to physical objects or substances (e.g., bottles, tablets, or chemicals). The shaded boxes in Figure 1 indicate that the *generic ingredient set* and *clinical drug* levels of description are pure abstractions that clinicians may find useful in thinking about drugs.

The top level in Figure 1 represents packaged drug products, which are found in a pharmacy's inventory and consist of manufactured formulations (e.g., tablets, capsules, solutions, etc.) that can be packaged in a variety of ways (e.g., ampuls, vials, bottles, blister packs, etc.). Special packaging, such as dispenser packs for triphasic oral contraceptives, may be used to facilitate appropriate therapeutic use. A number of attributes, such as the size and cost of the package, are relevant only at this level of description.

1091—8280/98/$5.00

The bottom level in Figure 1 represents ingredients, i.e., the chemicals or substances that are combined into a manufactured formulation. An ingredient may consist of different components, such as a "base ingredient" and a salt. (For example, the ingredient "chloroquine phosphate" consists of the base ingredient "chloroquine" plus the "phosphate" salt.)

Within a given manufactured formulation, each ingredient is present in a particular quantity or strength. Ingredients having therapeutic intent are referred to as "therapeutically active ingredients" in the figure, while other ingredients are referred to as "inactive." However, it should be remembered that so-called "inactive" ingredients can also have biological effects. For example, a lactose-intolerant patient may have an adverse reaction to tablets that use this "inactive" substance as an excipient.

Two abstract levels of description are interposed between ingredients and manufactured formulations in Figure 1. These abstractions are useful primarily as grouping mechanisms. A *generic ingredient set* groups together all manufactured formulations that contain a particular set of therapeutically active ingredients, which can be referred to by a particular generic name (e.g., "trimethoprim/sulfamethoxazole" or "diazepam"), irrespective of the dose form or the strengths of the individual ingredients. This level of description can be useful in representing drug-drug interactions.

Another clinically useful abstraction groups together all manufactured formulations having the same set of therapeutically active ingredients and associated strengths within the same dose form (e.g., "any ampicillin 250 mg capsule"). This abstraction is especially useful to physicians writing drug orders and to dispensing pharmacists considering therapeutic substitution.

AN UNDERLYING INFORMATION MODEL

In order to represent drugs at different levels of description in operational computer systems, a formal information model must be constructed to define the relevant descriptive levels and the relationships among them. Over the years, vendors of commercial drug databases have developed information models focused on target markets that are primarily concerned with certain levels of description. An effort to develop an information model that encompasses all of the descriptive levels mentioned in Figure 1 has recently been reported[3].

An adequate information model must be able to represent the complex relationships that can exist between a drug at *any level of description* and other information that may be of clinical interest. The modeling issues entailed in representing dose forms, routes of administration, drug classification schemes, and medical problems (in the contexts of indications, contraindications, or adverse effects) deserve special

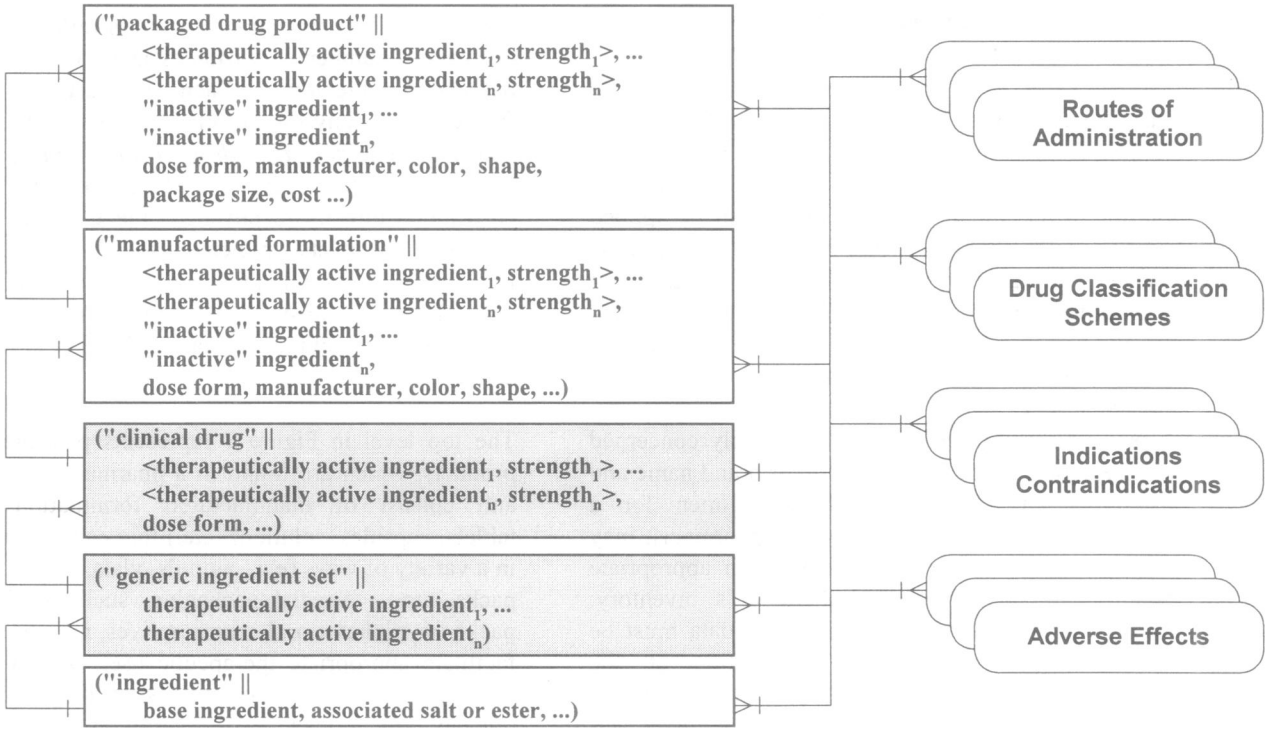

Figure 1: Levels of Drug Description

comment. Although the five levels of description shown on the left side of Figure 1 are depicted as a strict hierarchy, the right side of the figure is an attempt to illustrate that drugs described at any of these levels can participate in relationships to other information (e.g., drug classification schemes) that are poly-hierarchical or many-to-many in nature.

Dose Forms

In Figure 1, *dose form* is depicted as an attribute at the *packaged drug product*, *manufactured formulation*, and *clinical drug* levels of description. A dose form can be defined as the physical state of a manufactured formulation. There are many basic dose forms which deliver drugs to specific sites of administration. Some common examples are tablets, capsules, ointments, solutions, suspensions, and suppositories.

Manufacturers may modify these basic formulation types to improve the delivery of specific drugs. For example, a "12 hour sustained release" capsule incorporates particles with differing dissolution rates such that the drug is released over 12 hours.

An additional consideration is that the dose form ultimately administered to a patient may have a different physical state from the dose form produced by the pharmaceutical manufacturer. Many drug products are manufactured as lyophyllized powders, which must be reconstituted with an appropriate diluent to form a solution or suspension before being administered to a patient. The amount of diluent used will obviously affect the concentration of the therapeutically active ingredients in the substance that the patient ultimately receives, and different diluents may be required in different clinical situations.

Routes of Administration

In order to completely specify how a drug can be "given to the patient," we must consider both the *site* and the *method* of administration. For example, there are many different methods of administering a drug intravenously, including "IV bolus," "IV piggyback," and continuous infusion.

Deciding upon the most appropriate representation for routes of administration within an information model is somewhat problematic, since routes can be related both to indications and dose forms. These relationships are evident in the drug approval process, during which the U.S. Food and Drug Administration approves specific manufactured formulations of a drug for use in specific clinical indications by specific routes of administration. Moreover, there can be an implicit relationship between dose form and route, in that a particular dose form can be associated with only certain sites of administration regardless of the indication.

An information model should support multiple views of routes of administration as they relate to drugs at various levels of description. One approach is to link a set of routes, irrespective of clinical indications, to the *clinical drug* level. Another approach would represent individual routes only in association with specific clinical indications. A third approach is to link routes of administration to the *generic ingredient set* level, irrespective of dose form. Such a "routed generic" representation is useful to prescribing physicians when, for example, they only care that a patient receives a particular dose of oral medication — regardless of whether the medication is dispensed in the form of a tablet, capsule, or suspension.

Drug Classification Schemes

Drugs can be classified on the basis of different considerations, including receptor pharmacology (e.g., beta-adrenergic blocker or serotonin-specific re-uptake inhibitor), chemical structure (e.g., benzodiazepine, aminoglycoside, or quinolone), and therapeutic use (e.g., antihypertensive, sedative, or diuretic). Propranolol, for example, can be classified as a beta-adrenergic blocker, an antihypertensive, an anti-anginal, and an anti-migraine agent. Many classification schemes are available, and some, such as the BNF[4], are mandated in certain countries. Thus, an information model must be able to support multiple classification hierarchies for drugs at whatever level of description may be appropriate.

Medical Problems

The relationships between drugs and medical problems are central to many clinical decision support applications. The difficult issues entailed in encoding medical problems using controlled vocabularies are well known[5,6,7]. Although the information model being discussed need not attempt to address these issues in a general sense, it does need to handle issues that may arise in the context of representing drug indications.

A single drug may have multiple indications that require different dosing regimens and routes of administration. An indication may be expressed in rather precise language, such as "prophylaxis of cytomegalovirus retinitis in immunocompromised

patients," reflecting the clinical research studies that were performed during drug development. Since a single term from a controlled biomedical vocabulary may not adequately represent a drug indication in a manner that is "true to the labeling," some sort of compositional scheme may be needed. Moreover, drug manufacturers may not seek regulatory approval for legitimate new uses of marketed products, and the approved indications for the same drug may differ from country to country. The maintainers of a drug knowledge base must decide whether or not to include a particular unlabeled indication, based on their assessment of the evidence supporting this off-label use in the medical literature. In any event, decision support applications will often require information about why the drug is being prescribed if they are to provide patient-specific advice. Even when medical problems are expressed in a controlled vocabulary, such applications need strategies for obtaining and disambiguating indication information for a specific patient, perhaps through interaction with an electronic medical record system and/or a clinical user.

NAMING CONSIDERATIONS IN A CONCEPT-BASED DRUG VOCABULARY

Concept-based vocabularies must typically address the issues of *granularity*, *synonymy*, and *ambiguity*[8]. As we have proposed[3], the foundation of an underlying information model should be a highly granular representation from which multiple views can be derived.

The issues of synonymy and ambiguity must be considered as they apply to the *names* of drugs at various levels of description. In its most basic form, a concept in a controlled vocabulary simply consists of a unique identifier associated with a list of terms that share the same meaning (in the judgment of human reviewers). A concept-oriented representation is desirable because different character strings can have the same meaning (synonymy) and the same character string can have different meanings (ambiguity). These issues can be handled by using unique concept identifiers for key elements of the information model, such as drug descriptions at each level, dose forms, and routes of administration.

Assigning names to drug concepts at different levels of description is relatively straightforward for drugs having only a single therapeutically active ingredient. This task is considerably more complex, however, for combination drug products, which contain multiple therapeutically active ingredients.

For single-ingredient drugs, the generic name (e.g., "diazepam") is entirely adequate for concepts at the *ingredient* and *generic ingredient set* levels. When synonyms exist within a concept, one of these synonyms could be designated as the preferred name for the concept in a given country. For example, "acetaminophen" in the U.S. is known as "paracetamol" in the United Kingdom, while "albuterol" in the U.S. is called "salbutamol" in Canada. Names for concepts at the *clinical drug* level would have to include the strength and dose form in addition to the generic name (e.g., "diazepam 10 mg tablet"). Concept names at the *manufactured formulation* level would include trade names, strengths, and dose forms (e.g., "Valium 10 mg tablet"). Thus, "Tylenol 325 mg tablet" and "Advil 325 mg tablet" would be separate concepts at the *manufactured formulation* level, but both of these concepts would be linked to a single concept at the *clinical drug* level (i.e., "acetaminophen 325 mg tablet"). Concept names at the *packaged drug product* level would reflect package size (e.g., "Valium 10 mg tablets, bottle of 100"). Concepts at this level of description would be used only where package size is relevant, such as in pharmacy inventory control applications. While the concept names at some levels of description reflect combinations of different properties (such as generic or trade name, strength, dose form, and package size), an underlying database representation should have separate fields for each of these properties.

For combination drug products, the most complex naming problems arise at the *generic ingredient set* and *clinical drug* levels, where a concept name would consist of a list of generic names for therapeutically active ingredients. Although the concept names for combination drugs at these levels would be rather unwieldy (especially for products such as multivitamins or cough and cold preparations), the corresponding concept names at the *manufactured formulation* level would be more succinct. For example, the brand name "Robitussin®-CF," which contains implicit strength and dose form information, would be an appropriate concept name at the *manufactured formulation* level. Corresponding concept names at the *clinical drug and generic ingredient set* levels are shown below:

Example *Clinical Drug* Concept Name

guaifenesin 100 mg/
phenylpropanolamine hydrochloride 12.5 mg/
dextromethorphan hydrobromide 10 mg in each 5 mL

Example *Generic Ingredient Set* Concept Name

guaifenesin/
phenylpropanolamine hydrochloride/
dextromethorphan hydrobromide

Because drug combinations typically do not have official generic names, cumbersome concept names may be unavoidable if we want to name the abstractions that enable us to group together similar products from different manufacturers. In order to ensure a consistent naming convention, a sequence number could be assigned in an underlying database for each ingredient within a concept at the *clinical drug* or *generic ingredient set* levels. Such a compositional grammar could be applied either by an algorithm or by an editorial policy (perhaps based on a particular drug classification scheme) enforced by human editors. Nevertheless, application designers would probably want to refer to combination drugs by their more lexically succinct concept names at the *manufactured formulation* level, just as clinicians usually do in their everyday conversations.

IMPLEMENTATION, STANDARDS, AND MAINTENANCE ISSUES

The conceptual model that has been described underlies an ongoing development effort at First DataBank for a next-generation product known as the *Multilex Drug Knowledge Base*. Major features of this model reflect lessons learned over many years of collecting data for the company's current domestic and international drug data files. These information sources are intended for use by a wide variety of system vendors, including providers of retail pharmacy, electronic medical record, and pharmacy benefit management systems. Empirical validation of the new model can be approached in three phases. First, the adequacy of the five basic levels of description in the model can be assessed by populating it with a large number of drugs. As clinical information about these drugs is added to the model in the next phase, the utility of linking this information drugs at the most appropriate level of description can be assessed. In the third phase, the experiences of clinical system vendors and end-users with the fully populated knowledge base will provide the best opportunity for empirical validation of the model.

This conceptual model has been submitted to HL7[9], which is currently investigating strategies to enhance inter-operability among health care information systems by addressing drug-related vocabulary issues. Existing biomedical vocabularies do not have drugs as their primary focus and consequently do not address the representational issues that have been discussed in this paper. Population and maintenance of a drug knowledge base that implements this conceptual model is a resource-intensive task, requiring the full-time efforts of a staff of clinical pharmacists and other professionals. Frequent updates are needed to keep the content current, especially at the *packaged drug product* and *manufactured formulation* levels. The degree to which such content should be available in the public domain, perhaps by including it in the UMLS Metathesaurus[10], is currently a topic of active debate within the informatics community. Regardless of the outcome of that debate, it is hoped that the conceptual model described in this paper can make a worthwhile contribution to an emerging set of enabling infrastructures for health care informatics.

References

[1] Blois MS. Medicine and the Nature of Vertical Reasoning. N Engl J Med 1988;318:847-851.

[2] Blois MS, *Information and Medicine: The Nature of Medical Descriptions*. Berkeley: University of California Press, 1984.

[3] Broverman C, Kapusnik-Uner J, Shalaby J, Sperzel D. A Concept-Based Medication Vocabulary: An Essential Element for Pharmacy Decision Support. PPMQ 1998;18:1-20.

[4] *British National Formulary*. London: British Medical Association, 1997.

[5] McDonald CJ. The barriers to electronic medical record systems and how to overcome them. J Am Med Inform Assoc 1997 May;4(3):213-221.

[6] Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures. J Am Med Inform Assoc 1997 May;4(3):238-251.

[7] Cimino JJ. Review paper: coding systems in health care. Methods Inf Med 1996 Dec;35(4-5):273-284.

[8] Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: representing different views of biomedical concepts. Bull Med Libr Assoc 1993 Apr;81(2):217-222.

[9] Health Level Seven. Ann Arbor, Michigan, Version 2.3, 1996.

[10] Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998 Jan;5(1):1-11.