Assessing the Impact of HL7/FDA Structured Product Label (SPL) Content for Medication Knowledge Management

Gunther Schadow MD, PhD

Regenstrief Institute and Indiana University School of Informatics, Indianapolis, IN

The amount and quality of the SPL drug knowledge which has been released so far is assessed. All published labels were loaded into a relational database and classified to create vendor-independent descriptions. While SPL labels cover only 23% of RxNorm clinical drugs, they still describe 78% of actual community pharmacy dispenses records. SPL descriptions agree well with RxNorm. SPL can be used as the primary source of drug information for e-prescribing systems once the upcoming FDA listing rule takes effect. In the interim, existing gaps can be temporarily closed with RxNorm or other sources.

Introduction

With an increased broader use of Electronic Health Records (EHR) and Computerized Provider Order Entry (CPOE) systems (particularly e-prescribing systems), high quality drug knowledge in computer processable form is urgently needed. Traditionally, producers of large EHR and CPOE systems rely on 3 kinds of sources for drug knowledge: (1) Commercial database products such as First Data Bank (FDB), Micromedex, and Cerner/Multum; (2) public sources, such as the Veterans Administration’s (VA) NDF-RT, and the National Library of Medicine’s (NLM) RxNorm; and (3) in-house created knowledge such as the Regenstrief Medical Gopher CPOE system.

All these knowledge sources have in common that they (a) involve laborious processes to excerpt, encode, compile and reconcile data from various primary sources, and (b) represent this data in a variety of proprietary formats, different conceptual models, using different terminology, and are therefore not interoperable and difficult to integrate.

In the U.S., the pharmaceutical industry and the Food and Drug Administration (FDA) together are the authoritative source for drug information. The Physician’s Desk Reference (PDR), the reference book most-frequently consulted by American Physicin’s, simply compiles all the labels of currently marketed drugs approved by the FDA. The FDA’s drug listing service registers marketed drugs and administers the National Drug Codes (NDC), which is the most widely used drug-nomenclature in the U.S.

The FDA has embarked on an initiative to improve drug knowledge appropriate for use in clinical information systems. This includes the electronic labeling guidance requiring that drug labels be submitted electronically, the Physicians Labeling Rule (PLR) that mandates more user-friendly labels. The new proposed drug listing rule seeks to close the many gaps of the FDA’s drug listing databases by finally taking full control of the process of NDC code assignments. These initiatives use the Health Level-7 (HL7) version 3 Structured Product Labeling (SPL) standard for representing human readable label documents with computer-processable drug knowledge.

Since 2006, pharmaceutical manufacturers produce SPL labels for their products and FDA has released an increasing number of these labels through NLM’s DailyMed [http://dailymed.nlm.nih.gov] As of March 2007, 2273 labels are available, covering approximately 78% of prescription-drugs currently marketed by their original innovators. Coverage of generics is less, and repackaged and over-the-counter (OTC) drugs are not included at this time. These gaps will be closed with the upcoming listing rule, and all legally traded medicines will be covered by SPL labels.

To understand the impact of SPL labels for current drug knowledge management and CPOE system implementation, this paper investigates (1) if SPL labels are sufficient as an exclusive source for drug information for e-prescribing systems today; (2) if SPL labels can be used directly in conjunction with other knowledge sources. Since labels describe specific manufactured products but clinicians refer to more abstract drug concepts, we investigate (3) if it is possible to use SPL as a source for abstract drug knowledge, and (4) if SPL content agrees with other drug knowledge sources.

Methods

All 2217 labels available on DailyMed were downloaded as of February 23, 2007 and fed into an open-source HL7 v3 based data-infrastructure (HL7 JavaSIG). The HL7 Reference Information Model (RIM) is used directly as the object model. The HL7 RIM represents medicines, packages and ingredient-substances as (physical) Entities related to each other through Roles which specify the type of relationship, e.g., active or inactive ingredient, active moiety and container-content. Regardless of the type, all roles specify the amount of an ingredient over the amount of medicine, or the amount of medicine in a package as a
single computable quantity ratio. The HL7 Java SIG software can parse any HL7 v3 XML document, including SPL labels, and integrate them into a relational database using the Hibernate object-relational persistence. The relational database schema straightforwardly implements the HL7 RIM model which makes it easy to analyze all the SPL description with standard SQL. Through our system all SPL labels are truly integrated, for example, the objects representing ingredient substances are instantiated only once per each substance across all SPL labels and all drugs with this substance refer to that same identical object.

### Measuring Coverage

For a realistic measure of coverage of current SPL content, the NDC codes in SPL content were matched with NDC codes referenced in RxNorm as well as NDC codes cited in a convenience sample of 89748 NDC codes from community pharmacy dispense records between June 2004 and February 2007. To compare, all NDC codes were converted into the 11-digit format. Realizing that SPL labels do not cover generic or OTC drugs very well, simply comparing NDC codes yields a meaningless measure of coverage. Instead, coverage was compared using RxNorm’s “Semantic Clinical Drug” (SCD) abstraction to actually measure the coverage of substance combinations independent of their manufacturers.

### Classifying Medicines

The drug descriptions in SPL are entirely product and package specific but SPL labels could only be used in e-prescribing systems, if such systems allow clinicians to refer to drugs more generally and thus need to abstract from the product and manufacturer specific details in SPL. Figure 1 shows several possible abstractions. The “Semantic Clinical Drug” (SCD) of RxNorm, is a combination of abstract ingredients (roughly equal to “active moiety” defined by FDA rules) and dose form. RxNorm never considers route of administration as a separate property. In the Regenstrief CPOE system, an abstraction by moiety, route and abstract form (with only few distinctions “solid” vs. “liquid”) has been found most useful.

To prove that SPL labels can be abstracted to more general descriptions, we classified the SPL descriptions using no other knowledge and using only SQL. The systematic approach is described here in some detail because it is simple and generally applicable to any categorical classification problem. (Previously the author has used this same approach for bulk patient matching). Classification is based on a symmetric, reflexive, and transitive relation (equivalence relation) between pairs of objects of the same class. In such a relation, one can choose any member of the class to represent the entire class, and then collect all objects equivalent to this representative as the members of the class. Because the relation is reflexive, the representative of the class is itself a member of the class.

For classifying equal combinations of ingredients (or any “traits”), the initial equivalence relation is built by relating pairs of candidate members by joining on

```sql
WITH ObjectTrait AS (  
  SELECT object1Id, traitValue FROM ...
), ObjectTraitCount AS (  
  SELECT object1Id,  
    COUNT(1) AS traitCount  
  FROM ObjectTrait  
  GROUP BY object1Id
), ObjectTraitWithCount AS (  
  SELECT * FROM ObjectTrait  
  INNER JOIN ObjectTraitCount  
  USING(object1Id)
), EquivalenceRelation AS (  
  SELECT s1.objectId AS object1Id,  
    s2.objectId AS object2Id  
  FROM ObjectTraitWithCount s1  
  INNER JOIN ObjectTraitWithCount s2  
  USING(traitValue, traitCount)  
  GROUP BY s1.objectId, s2.objectId  
  HAVING COUNT(1) = MAX(traitCount)
)  
SELECT object1Id AS memberId,  
  MIN(object2Id) AS classId  
FROM EquivalenceRelation  
GROUP BY object1Id

Exhibit 1: SQL query for classifying objects (e.g., drugs) based on complete agreement on common traits (e.g., ingredients). The class is identified by the id of one of its members.

![Figure 1: Examples of useful abstractions of medicine terms. We can abstract from any property, packaging (P), brand (B), dose form (F), route (R), strength (S), active-ingredient (I), moiety (M). The signature of letters (e.g. “MisRF”) marks the properties differentiated (upper-case) vs. abstracted (lower-case). Examples: “Oral penicillin” (MisR), “Penicillin as oral liquid” (MisRF). This shows clearly that abstractions of drug terms do not fit into a single hierarchy.](image)

---

**Figure 1:** Examples of useful abstractions of medicine terms. We can abstract from any property, packaging (P), brand (B), dose form (F), route (R), strength (S), active-ingredient (I), moiety (M). The signature of letters (e.g. “MisRF”) marks the properties differentiated (upper-case) vs. abstracted (lower-case). Examples: “Oral penicillin” (MisR), “Penicillin as oral liquid” (MisRF). This shows clearly that abstractions of drug terms do not fit into a single hierarchy.
their common traits, and selecting from this candidate relation only those pairs which have all their traits matched; i.e., such pairs of matched traits are included in the candidate-set exactly as many times as each candidate members’ total number of traits. The complete query is shown in Exhibit 1.

Once the initial classes were formed by ingredient and moiety, with or without consideration of strength, sub-classes were discovered according to the classical Aristotelian principle by common generalization and specific difference. Exhibit 2 shows all steps up to the definition of the equivalence-relation at which point the classification proceeds as in Exhibit 1 above. Because in this approach classes are represented by one arbitrary member, one can easily join multiple levels of the classifications thus constructing any desired “hierarchy” using only the simple SQL queries shown.

Comparing Drug Description Details
To validate the classification described above and to expose any errors in SPL or RxNorm, SPL and RxNorm descriptions of common NDCs were compared for agreement in their ingredients. This would reveal if some ingredients have been left out in one but included in the other. The FDA ingredients and moieties (coded in SPL using the FDA’s Unique Ingredient Identifier, UNII) were mapped to RxNorm using the synonyms which RxNorm inherits from the UMLS. Thus, 1278 of 1327 ingredients (96.3%) were mapped by complete name, leaving only 49 to map manually and completely, albeit in some cases not uniquely.

Results

Measuring Coverage
SPL covers only 3% of the NDC codes known to RxNorm, however, SPL covers 2630 (23%) of the 11476 RxNorm “Semantic Clinical Drugs” (SCD) and all SPL drugs had a matching SCD. Of the NDCs in the community pharmacy dispense records 97% had a matching SCD (94% of unique NDCs). In the dispense records, 69062 of 88914 occurrences of SCDs (78%) were covered by SPL (1271 of 2052, 61.9% of unique SCDs). Among the SCDs that were dispensed at least 100 times, SPL coverage was 80% and reached a maximum of 84% for drugs dispensed more than 200 times (approximately 46% of occurrences).

Classifying Medicines
Using the classification algorithm described above, the 12 reasonable of the 16 possible variants of classes described in Figure 1 were created. As Table 1 shows, the number of classes discovered range between 950 at the highest abstraction level (Misrf = active moiety-combination only) to 3056 at the lowest. All class-membership relations had the same cardinality, i.e., the classifications always included all the available drugs.

When classifying RxNorm by active moiety, we found that of 2626 RxNorm SCD descriptions 99% were classified under a single class and only 28 (1%) were associated with 2 classes. This indicates good agreement between our approach of classifying SPL descriptions and RxNorm’s classes.

Comparing Detail Drug Descriptions
The descriptions of 7083 NDC codes in SPL 6476 (91%) agree exactly with RxNorm in their ingredient moieties. Review of the remainder revealed a mix of reasons for disagreement in nuances; however, all of these differences were consistent with agreement. For example, some drugs have two ingredients, both with the same active moiety and not differentiated in RxNorm. Others were due to slight imprecision in the ingredient mapping. None of them were due to a blatant omission or falsely specified ingredients.

```
WITH ClassMemberDifference AS ( 
    SELECT c.classId AS genusId, 
        c.memberId AS objectId, 
        p.value AS difference 
    FROM BroaderClassMember c 
        INNER JOIN DifferentiatingProperty p 
    USING (memberId) 
), EquivalenceRelation AS ( 
    SELECT m1.objectId AS object1Id, 
        m2.objectId AS object2Id, 
        genusId, difference 
    FROM ClassMemberDifference m1 
        INNER JOIN ClassMemberDifference m2 
    USING (genusId, difference) 
) 

Exhibit 2: SQL query for refining categorical classification by common generalization and specific difference.
```
Discussion

Looking at only the miniscule 3% of coverage of raw NDC codes in SPL labels, it is easy to underestimate the importance of SPL label content for drug knowledge management today. This should be a motivation for FDA and all involved parties to expeditiously proceed with the finalization and implementation of a listing rule that will require listing data to be submitted in SPL form and will close the known deficiencies in the NDC assignment process. Presently, many repackaged medicines can be marketed with NDC codes unknown to the listing service. With the new listing rule, such medicines would be “misbranded”, out of compliance, and subject to seizure by the FDA. This will be a driving force for a much needed complete and up-to-date public knowledge base of medications.

However, even today, SPL content covers 78% of the mass of actually prescribed drugs, which affirms that it is reasonable to proceed now with the implementation of SPL knowledge content in e-prescribing and EHR systems. For a transitional period, however, SPL content needs to be augmented using other, more complete sources, such as RxNorm. As SPL coverage becomes more complete in the next 2 years these supplementing sources can slowly be superseded.

It is not difficult to cast RxNorm content into HL7 SPL data structures (Figure 2) and to merge the RxNorm content with current SPL content from DailyMed on that common ground. The relationship between RxNorm and the HL7 RIM model for drug descriptions is straight-forward: RxNorm’s “Semantic Branded Drug” (SBD) corresponds to the SPL Medicine class, the SBD Component (SBDC) corresponds to SPL’s ActiveIngredientRole, and RxNorm’s Ingredient (IN) corresponds to SPL’s Substance and ActiveMoietyEntity.

A slight advantage of the SPL format may be that navigating the HL7 RIM-based SPL model is easier than the UMLS-based RxNorm model. For example, to connect a medicine with its ingredient substance requires 4 joins in RxNorm from RxNCONSO (SBD) via RxNREL (consists_of), RxNCONSO (SCDC), and RxNREL (has_ingredient), to RxNCONSO (IN), whereas in SPL it is only 2 joins from Medicine (Entity) via ActiveIngredient (Role), to Substance (Entity).

The true advantage of SPL content is a more complete (and standardized) coverage of details of medications which can be of interest in EHR systems. For example, route, inactive ingredients (useful for allergy checking), other distinguishing features such as color, shape, and imprint of solid dose forms (help identify unknown medicines found with patients), and packaging details (relevant for pharmacy systems), are all described in SPL content, but not in RxNorm. In addition, packaged medicines are the true referents of NDC codes, which can be important when using drug dispense records for compliance monitoring.

The strategic advantage of using SPL today is that the SPL knowledge representation employs a conceptual model that is reused for many purposes including actual prescriptions in the HL7 standard. For example, the HL7 RIM-based SPL model contains a single consistent formalism for specifying quantities in relationships (whether ingredients or container-content relationships), whereas in RxNorm we had to parse this data from the name-string of the “Drug Compo-
We have shown here that one can effectively process HL7 v3 content using a combination of object-oriented software and relational databases which is readily available as open source. Finally, with FDA’s new “indexing” guidance, increasing amounts of clinically relevant information will be added into the SPL labels using a model and format which is already defined today. This will include annotation of medicines and substances with NDF-RT classes for mechanism of action, physiologic effect, and chemical structure, as well as indications, contraindications, adverse effects and maximum dose.

We also confirmed in this study that the most important piece of a drug knowledge base is the presence of high quality drug descriptions in as much detail as might possibly be needed. Based on these descriptions, anyone can easily create desired generalizations and “hierarchies” using only a few simple SQL queries. Therefore clients of drug knowledge sources need not be too concerned with such “hierarchies” as an intrinsic value and should instead demand complete and detailed descriptions.

Incidents such as the recent withdrawal of Vioxx from the market demonstrate that drug knowledge is subject to sudden change, and must therefore be kept up to date in a timely manner. Drug knowledge bases should be considered a product-catalog rather than common biomedical terminology that changes less frequently. SPL labels and the timely distribution process by which they appear on (and are withdrawn from) the DailyMed service are ideally suited to meet the need for detailed descriptions and timely updates as the authoritative parties themselves manage that content.

Present SPL labels were not free of technical errors, which is common for all standardized data interchange. In the course of this work, several systematic technical errors in the SPL encodings had been detected, such as missing or wrong code system identifiers, or other format errors. These were reported to the FDA, and using an XSLT-transform, these errors could be provisionally corrected. This work will help improve the SPL quality management process. However, apart from these technical errors, no significant content errors had been found in this study and there is excellent agreement of content between SPL labels and other drug data sources, such as RxNorm.

Conclusion
SPL content covers the majority of currently prescribed drugs in outpatient settings, and RxNorm can be used to bridge current coverage gaps of SPL. SPL has more detail on the drugs that are covered, but RxNorm and SPL have excellent agreement on the detail contained in both. The HL7 v3 SPL data is accessible with object-oriented software and relational database systems. User-defined abstraction from manufactured drugs can easily be created on the basis of detailed drug descriptions in the relational database.

Acknowledgements
This work was performed at the Regenstrief Institute and is funded in part by the Agency for Healthcare Research and Quality (AHRQ) grant R01 HS15377 and the Food and Drug Administration (FDA). Without Randy Levin’s genial leadership this work would not have been possible.

References
10 Schadow G, Gitterman S, Boyer S, Dolin RH eds. HL7 v3.0 structured product labeling, release 2 [standard]. Ann Arbor, MI, Health Level Seven, 2005
11 Levin R. Personal communications.
12 http://aurora.regenstrief.org/javasig
13 Schadow G, Russler DC, McDonald CJ. Conceptual Alignment of Electronic Health Record Data with Guidelines and Workflow Knowledge. Int J Med Inf. 2001 Dec;64(2-3):259-74