Managing terminology change for clinical medications in the Learning Health System

James R. Campbell, MD\textsuperscript{1}; Jay Pedersen, MA\textsuperscript{1}; Lee Peters, MS\textsuperscript{2}; W. Scott Campbell, PhD MBA\textsuperscript{1}; Olivier Bodenreider, MD PhD\textsuperscript{2}

\textsuperscript{1}University of Nebraska Medical Center, Omaha, NE; \textsuperscript{2}US National Library of Medicine, National Institutes of Health, Bethesda, MD

Abstract

While preparing ontology metadata for a reload of our i2b2 data warehouse, we discovered many EHR records with instances of RxNorm and NDC codes that could not be found in UMLS, in the current version of RxNorm or in the FDA National Drug Code Directory. Exploring the editorial policies of the FDA and NLM, we determined that RxNorm and NDC releases only reflect codes for drugs currently in use in the U.S. and are most useful for writing medication orders or reporting medication dispense events. We collaborated with colleagues at the NLM to identify services to provide RxNorm, NDC and VA drug classification codes and relationships that were historically comprehensive and useful for analytics in a longitudinal clinical data warehouse extracted from our electronic health record(EHR). We explain how those services may be employed by a site manager to organize and index the historical medication record data for warehousing. We further propose expanded editorial criteria for the RxNorm and NDC code sets to support the research data warehouse use case within the Learning Health System.

Introduction

For compliance with Meaningful Use, the Office of the National Coordinator for Healthcare Information Technology(ONC) has specified a set of standard terminologies to be deployed in electronic health records(EHR) in the US\textsuperscript{1}. In addition to HIPAA transaction term sets, SNOMED CT\textsuperscript{2} is to support conditions and problems, LOINC\textsuperscript{3} encodes laboratory and test data, RxNorm\textsuperscript{4} is mandated for medication orders and prescriptions. The National Drug Code(NDC), is a universal product identifier for human drugs in the United States mandated by the FDA. The code is present on all medication packages and inserts in the US and is employed as a transaction code standard for medication dispense and administration events\textsuperscript{5}.

EHR data in turn, are expected to be re-used within the Learning Health System(LHS)\textsuperscript{1,6} to support clinical decision making, clinical research and public health. Data networks supporting collaborative research and public health\textsuperscript{7,8} missions have sprung up in the last five years, heavily fueled with data extracted from the EHR. Structuring that extracted data is critical to interoperability and the ability to merge multiple EHR data sets. Terminologies mandated by ONC are therefore central to the architecture of LHS data networks. In order to support the complexities of interoperability the standards bodies managing those terminologies should comply with requirements established by National Committee on Vital and Health Statistics(NCVHS) fifteen years ago\textsuperscript{9}. One key feature of those requirements is support for graceful evolution for meaning of EHR data over time as the reference terminologies evolve by issuing new term references and retiring others.

With the announcement of Meaningful Use (MU), ONC offered incentives for deployment and use of EHRs in the US. Those incentives were fantastically successful and by 2015\textsuperscript{10} 78% of physician offices and 96% of US hospitals employed an EHR that is certified compliant with the functional and interoperation standards specified by ONC. At University of Nebraska, we converted our EHR to Epic\textsuperscript{®} in 2012 to assure that we could meet MU goals. Terminology standards for structuring EHR data are central to MU and certified systems such as Epic\textsuperscript{®} employ them in data dictionaries within their product, although that fact may not be obvious to the clinical user or the informatics specialist.

In this paper we recount the informatics issues we confronted while extracting and organizing our Epic\textsuperscript{®} EHR medication orders, dispensing and administration events into our i2b2\textsuperscript{13} data warehouse. We discuss editorial and publication strategies and documentation resources available for RxNorm and NDC. We report our collaborations with staff at the NLM and their development of resources to support i2b2 metadata build that maintains historical context for EHR data instances now encoded with retired or obsolete concepts. Finally we offer suggestions for enhancing temporal management of these terminologies to comply with the requirements identified by NCVHS and to simplify medication data re-use for other healthcare institutions.
Background

FDA National Drug Code Directory

The FDA National Drug Code (NDC) Directory provides “a current list of all drugs manufactured, prepared, propagated, compounded, or processed by it for commercial distribution,” in which “Drug products are identified and reported using a unique, three-segment number, called the National Drug Code (NDC), which serves as a universal product identifier for drugs”. The NDC Directory is updated daily.

In the NDC Directory 2-segment codes are used to identify the drug products, while 3-segment NDCs (manufacturer-drug-packaging) identify drug packages. NDCs found in claims data and EHR data are generally standardized HIPAA format 11-digit codes derived from the 3-segment NDCs by padding missing digits with zeroes.

RxNorm

RxNorm is a standardized nomenclature for medications produced and maintained by the U.S. National Library of Medicine (NLM) in cooperation with proprietary vendors. RxNorm concepts are linked by NLM to multiple drug identifiers for each of the commercially available drug databases referenced in the UMLS® Metathesaurus®. In addition to integrating names from existing drug vocabularies, RxNorm creates standard names for clinical drugs.

RxNorm curates NDC codes derived from two terminologies – DailyMed and First Data Bank. Other source vocabularies also contribute NDCs to RxNorm, however those additional NDCs are not edited or managed by NLM. Nor does RxNorm include NDCs from the FDA National Drug Code Directory, as it was observed that DailyMed and First Data Bank better reflect drugs currently marketed in the US.

RxNorm is updated monthly and each version only contains active NDCs reflecting the current market state of drug products. With each new version of RxNorm, NDCs may be added or removed. RxNorm does not keep publish or archive obsolete NDCs.

The RxNorm API provides functionality to access the RxNorm dataset, including mapping from current NDCs to obtain the RxNorm concept identifier (RxCUI). Only active NDCs are mapped to RxNorm concepts. Conversely, the RxNorm API also associates an active RxCUI with any NDC, active or obsolete, ever curated by RxNorm. The RxNorm browser, RxNav, only provides access to active drug codes in RxNorm.

The UMLS® Metathesaurus® publishes content reported in RxNorm only every six months and reflects active RxCUIs from RxNorm and NDCs but does not support historical or archived terminology.

University of Nebraska i2b2 clinical data warehouse

I2b2 employs a star schema relational data model with keys which include concept codes linked to reference terminologies such as RxNorm or NDC. The user interface employs a metadata taxonomy which supports browsing and query functionality. An exemplar from our medication taxonomy is shown in Figure 1. Top level of this metadata taxonomy is depicted as Veterans Administration drug classes - agreed to be most clinically organized of the top level classifications available by our research network colleagues. To support aggregation queries of observation facts that aligned with common use cases, we organize the metadata hierarchy as follows:

```
VA Drug class\RxNorm Ingredient or multi-ingredient\RxNorm clinical drug or pack\NDC package\n```

Extracting our Epic® EHR data to our i2b2 data warehouse in 2017, we documented rich data sets compliant with drug terminology and other ONC standards as summarized by the i2b2 coded data counts of Table 1. During the refresh, we noted inconsistencies in the sets of RxNorm and NDC codes between RxNorm downloads from NLM and those record instances documented in Epic. In 2015 we had brought the problem with dispense events to the attention of NLM staff and asked for assistance in obtaining a complete history of NDC codes. Staff at the NLM advised that we use the RESTful services of the RxNorm API to build our metadata. They developed a new service call ‘getAllHistoricalNDCs’ which returns a comprehensive history of NDC codes for a given clinical drug. Unfortunately, package naming could not be provided with that service.
Figure 1. i2b2 client showing medication metadata

Table 1. i2b2 Observation fact counts by ONC terminology standard and datatype

<table>
<thead>
<tr>
<th>Terminology standard</th>
<th>EHR record type</th>
<th>Fact records in i2b2 Patients = 2,020 (in thousands)</th>
<th>Earliest data record in i2b2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RXNORM</td>
<td>Prescriptions and Medication orders</td>
<td>11,583</td>
<td>2/13/2002</td>
</tr>
<tr>
<td>NDC</td>
<td>Medications dispensed or administered</td>
<td>67,522</td>
<td>5/13/2009</td>
</tr>
<tr>
<td>ICD-10-CM</td>
<td>Encounter diagnoses</td>
<td>12,363</td>
<td>10/1/2015</td>
</tr>
<tr>
<td>SNOMED CT</td>
<td>Problems/Conditions</td>
<td>4,985</td>
<td>1/2/1984</td>
</tr>
<tr>
<td>CPT-4</td>
<td>Professional services</td>
<td>10,419</td>
<td>1/1/2011</td>
</tr>
<tr>
<td>ICD-10-PCS</td>
<td>Hospital procedures</td>
<td>74</td>
<td>5/3/2014</td>
</tr>
<tr>
<td>SNOMED CT</td>
<td>Procedure history</td>
<td>145</td>
<td>1/1/1992</td>
</tr>
<tr>
<td>LOINC</td>
<td>Laboratory results</td>
<td>206,857</td>
<td>1/1/1995</td>
</tr>
<tr>
<td>LOINC</td>
<td>Patient findings</td>
<td>69,267</td>
<td>1/15/1994</td>
</tr>
</tbody>
</table>
Examining the code sets between 2015 and 2017, we further realized that we were missing historical RxNorm concepts as well. In February 2017, we noted that 3771 RxCUIs were not in the release set and the RXNAV browser reported that these were either retired or obsolete and replaced. No history file or RxCUI change management transaction log was supported by NLM at that time.

Problem statement
The LHS requires that data accumulated in the course of clinical care be available for re-use and particularly be available to serve research projects such as those we are managing with i2b2. Terminologies featured within the ONC interoperability framework must therefore support long-term and historical data management. In the case of RxNorm and NDC however, the use cases supported are limited to current active treatment and drug manufacturing. Changing hats from clinician to data warehouse manager, we find that valid clinical data in our EHR is no longer viable due to lack of support for historical change documentation by the NLM and FDA. Lack of compliance with vocabulary management principles established by NCVHS poses a threat to re-use of medication data within the LHS.

Methods and Results
Development of RxNorm and NDC history services
NLM staff who are co-authors on this paper have been responsive to our issues as we have prepared the software to build/refresh our i2b2 metadata. Based upon limitations in temporal management we identified in UMLS and FDA downloads, we have built our protocols entirely upon RxClass and RxNorm service calls. NLM staff have built an API specifically for NDC code history that unfortunately does not have the package names we can obtain from the FDA site. For RxNorm codes that are retired or obsolete, they have collated transactions over the release history of RxNorm and have provided us support for our current build. They indicate that a service call for RxCUI history is under development.

I2b2Metadata build for medications
The procedure for creating i2b2 medication metadata is a multi-step process resulting in the creation of metadata rows whose overall structure is a mathematical tree, representing the visual folder structure associated with this metadata. At the root of the tree is the VA Drug Classification. An RxClass service call (Table 2, step 1) returns the VA Drug Classifications in a nested tree structure containing a descriptive name and a national Drug File – Reference Terminology(NDF-RT) code for each node. We traverse the tree and use its structure to build the first several levels of metadata.

We determine the RxCUI codes of generic drug products associated with each VA class hierarchy member by using an RxClass service that returns zero or more RxCUI codes of generic drug products associated with a VA Class (Table 2, step 2).

We identify the ingredients associated with those generic drug products by an RxNorm service class call that returns one or more RxCUI codes associated with the specified drug product (Table 2, step 3).

We retrieve the branded drug products associated with generic drug products by an RxNorm service class call that returns zero or more RxCUI codes of branded products given the RxCUI code of a generic drug product (Table 2, step 4).

Currently lacking a service call for historical RxCUIs, we then process an experimental history file of RxNorm codes provided by NLM staff developed from exhaustive compilation of past RxNorm releases. The historical information for each RxCUI code includes an associated ingredient RxCUI and a set of zero or more replacement RxCUI codes. A code with no replacement was retired; otherwise, it is an obsolete code. We iterate across the historical RxCUI codes, creating one or more metadata rows for each. We create a single metadata row for retired codes. For obsolete codes, we create a folder for the obsolete code with multiple leaves – one for the obsolete code itself and one for each replacement code. The rows are children of the associated ingredient in the VA Drug Hierarchy when the ingredient RxCUI exists in that hierarchy. Otherwise we insert rows that are children of the MISCELLANEOUS AGENTS folder of the VA Drug Hierarchy (NDF-RT code N0000029353).
Finally, we identify all NDC codes which have ever been manufactured to vend the clinical drug or pack. An RxNorm service call (Table 2, step 5) returns zero or more NDC codes for a given prescription RxCUI code. This includes a complete history of NDC codes ever in use. Each result becomes a metadata row, a child of the prescription drug. For those NDC codes in active use, we retrieve the reference files from the FDA download site\(^1\) and construct a description for the manufactured drug pack. For all other NDCs, we create a description using only the clinical drug description from RxNorm.

**Table 2. RxClass and RxNorm service calls to build i2b2 metadata**

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Variables</th>
<th>Service Call</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1 Obtain the list of all VA classes</td>
<td>None</td>
<td><a href="https://rxnav.nlm.nih.gov/REST/rxclass/classTree.json?classId=N0000010574">https://rxnav.nlm.nih.gov/REST/rxclass/classTree.json?classId=N0000010574</a> returns all VA Classes e.g., QUINOLONES (N0000183553)</td>
</tr>
<tr>
<td>Step 2 Obtain generic drug products associated with a VA drug class.</td>
<td>&lt;NDF-RT code&gt;, NDF-RT code associated with the VA class</td>
<td><a href="https://rxnav.nlm.nih.gov/REST/rxclass/classMembers.json?classId=%3CNDF-RT_code%3E&amp;relaSource=NDFRT&amp;rela=has_VAClass&amp;ttys=SCD+GPCK">https://rxnav.nlm.nih.gov/REST/rxclass/classMembers.json?classId=&lt;NDF-RT_code&gt;&amp;relaSource=NDFRT&amp;rela=has_VAClass&amp;ttys=SCD+GPCK</a> Example for: QUINOLONES (N0000183553) <a href="https://rxnav.nlm.nih.gov/REST/rxclass/classMembers.json?classId=N0000183553&amp;relaSource=NDFRT&amp;rela=has_VAClass&amp;ttys=SCD+GPCK">https://rxnav.nlm.nih.gov/REST/rxclass/classMembers.json?classId=N0000183553&amp;relaSource=NDFRT&amp;rela=has_VAClass&amp;ttys=SCD+GPCK</a> returns 46 distinct RxNorm generic drug products e.g., trovafloxacin 100 MG Oral Tablet (313522)</td>
</tr>
</tbody>
</table>

The February 2017 metadata build for i2b2 deploys 30,964 distinct historical RxCUIs and 455,583 distinct historical NDC codes.
Discussion

Use cases supported by the NLM and FDA for the RxNorm and NDC code sets at this time include predominantly the prescribing, ordering, dispensing and administering of medicinal substances currently being manufactured and distributed as part of the US pharmacopoeia. With the success of MU and widespread movement to repurpose EHR data as part of the LHS, these code sets must evolve editorially to serve the requirements of research and public health data sets. Recoding or mapping of these data to alternative classifications or terminologies is not viable except at the level of meaning asserted at the time of the original data record. Editorial shortfalls which should be corrected by NLM and the FDA to meet the use case requirements of the LHS include:

- Lack of history mechanism for reporting inception and retirement dates for RxCUI and NDC concept references
- Insufficient definitional material specifying the meaning and importance of a clinical drug or manufactured formulation that has been taken off the market
- Lack of structured history data for relationships of retired drugs or formulations to clinical, chemical and functional classification schemes
- NDC codes may be re-used by the manufacturer and are not unique and specific

NLM staff have started addressing some of these issues by providing access to some 400,000 obsolete NDC codes through an API service referenced above. An service in development will provide historical information for obsolete RxCUIs.

Speaking to the use case of implementing and updating an i2b2 or other data warehouse for medication events, the informatics specialist should be aware of the limitations and shortcomings of publicly available datasets from FDA and UMLS® - both Metathesaurus® and Terminology Server downloads.

More generally, the lack of support for obsolete drug codes has been noted as an issue for the analytics of observational datasets. For example, Homer et al. describe a complex pipeline combining a proprietary resource (Cerner Multum’s VantageRx) and the RxNorm API for linking NDCs from a large claims database to the Anatomical Therapeutic Chemical (ATC) drug classification system. Similarly, the Observational Health Data Sciences and Informatics (OHDSI) Consortium has also adopted RxNorm and NDC codes for representing medications in the OMOP Common data model, and have developed considerable terminology resources to manage current and obsolete codes.

Our experience developed in dialogue with NLM staff suggests that the RxClass and RxNorm APIs should be considered the sole most comprehensive source of clinical drug information for the US. We have published our procedures for building i2b2 medication metadata using the NLM service calls at http://github.com/SCILHS/scilhs-ontology.

Acknowledgements

This work was supported in part by NIH award U01 HG009455(Nebraska), Patient Centered Outcomes Research Institute(PCORI) award CDRN-1306-04631(Nebraska), Dr Ashok Mudgapalli and the Research Information Technology Office(Nebraska) and the Intramural Research Program of the NIH, National Library of Medicine.

References

1. Office of the National Coordinator for Health IT. Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap. 2015.
11. http://www.i2b2.org