Internship Report
AUDITING THE INTEGRATION OF RxNorm AND MEDICINAL PRODUCTS IN SNOMED CT

Jean Noël NIKIEMA
Bordeaux Population Health Center,
Lister Hill National Center for Biomedical Communications

U.S. National Library of Medicine, NIH
Mentor: Dr. Olivier Bodenreider

1 Background

Integrating RxNorm with the medicinal products in SNOMED CT generally contributes to increasing interoperability between drugs terminologies. In addition, this integration helps improve the description of medicinal products in both terminologies. In previous work, the strategy we used for this integration first consisted in comparing the definitional characteristics of clinical drugs in both terminologies and second in using the formal definitions of each clinical drug entity in RxNorm and SNOMED CT as the basis of this comparison. Our present work consisted in analyzing the results of this comparison in order to improve the algorithms and identify issues with the underlying data.

2 Material

We used the OWL file resulting from the integration of RxNorm (as of 02-Dec-2019) with SNOMED CT (preview of January 2020) performed by NLM staff.

3 Method

We performed three main steps to audit the integration process. (1) auditing of the mapping of definitional features, (2) categorizing the differences between inferred (induced by the integration process) and asserted (described in RxNorm) mappings, and (3) defining a course of action for each type of mismatch between the clinical drug entities involved in asserted mappings.

3.1 Auditing of the mapping of definitional features.

This step consisted in analyzing the mappings between ingredients and substances, and between dose forms and pharmaceutical dose forms (respectively in RxNorm and SNOMED

CT). As shown in Figure 1, we analyzed RxNorm ingredients according to their asserted mappings with substances in SNOMED CT:

Figure 1: auditing the ingredients mappings

- o 1-0 corresponds to ingredients without any mappings
- o 1-1 corresponds to ingredients mapped to only one substance
- o 1-N corresponds to ingredients mapped to multiple substances

We performed a lexical mapping to identify possible false negatives in 1-0 mappings (i.e., the substance exists in both RxNorm and SNOMED CT, but no mapping is asserted between them), and remove possible false positives in 1-N mappings (i.e., by selecting the closest match among asserted multiple mappings).

In addition, we identified RxNorm dose forms mapped to multiple SNOMED CT pharmaceutical dose form and unit of presentation pairs. In this case, we used general concept inclusion axioms (i.e., multiple equivalence axioms) to reflect the ambiguity of these mappings and maximize the chances of inferring mappings among clinical drugs.

3.2   Categorizing the differences between inferred and asserted mappings

This step consisted in determining the different types of errors leading to some differences between the inferred and asserted mappings among clinical drugs.

We explored the asserted mappings with no corresponding inferred mapping to find the characteristics of clinical drugs in RxNorm that went in the way of inferring the mapping (i.e., equivalence between the RxNorm and SNOMED CT logical definitions).

We first classified as erroneous mappings those mappings between one RxNorm clinical drug and multiple clinical drugs in SNOMED CT. These mappings should be reviewed manually for accuracy.

Second, we analyzed mapping errors according to the main definitional features (ingredients, dose forms), and the templates used for describing drugs. We analyzed remaining errors ("semantic errors") by identifying specific differences in the logical definitions of drug entities.

Figure 2: Categorizing the differences between inferred and asserted mappings

In practice, we annotated the clinical drugs in RxNorm according to errors in:
- Ingredients
    - o **SSU:** when the clinical drug contains unmapped ingredients in its description
    - o **ENM:** when the clinical drug contains erroneously unmapped ingredients
    - o **MBO:** when the basis of strength substance (BoSS) ingredient information is missing in the clinical drug description
    - o **MAI:** when the active ingredient information is missing in the clinical drug description
- Dose forms
    - o **SPE**: when the dose form in RxNorm is not mapped to a pharmaceutical dose form in SNOMED CT
- Template
    - o **ERR**: when the algorithm selects a different template compared to what is used in SNOMED
- Semantic:
    When all the definitional features of the clinical drugs in RxNorm, are correctly mapped, the remaining issues may be related to specific differences between RxNorm and SNOMED CT
    - o **BOS**: difference in the basis of strength substance(s) (BoSS)
    - o **VAL**: value difference
    - o **DFE**: dose form difference
    - o **SUM**: unit of measure difference

o **SAI**: difference in the active ingredient(s)

3.3   Defining a course of action

Based on the categorization of differences described above, we analyzed all RxNorm clinical drugs for which an equivalence in SNOMED CT is indicated in RxNorm (asserted mapping), but no equivalence is inferred based on the logical definitions ("asserted mapping only").

Figure 3: Course of action for improving the integration process

A specific course of action can be determined for each category of error.

4 Results

4.1   Auditing of the mapping of definitional features.

We identified 89 false positive ingredient-substance mappings, which we removed. We also identified 117 false negative ingredient-substance mappings, which we added. Finally, we identified 14 dose forms that are ambiguously mapped to SNOMED CT.

4.2   Categorizing the differences between inferred and asserted mappings

Table 1 describes the distribution of the RxNorm clinical drugs according to their asserted and inferred mappings with SNOMED CT clinical drugs. The inferred mappings were provided by the initial integration process. The SNOMED CT release contained 6,276 clinical drugs.

Table 1: Distribution of RxNorm clinical drugs according to the source of mapping before curation of the mapping of ingredients

		Asserted mappings		total
		Present	Absent	
Inferred mappings	Present	3,335	142	3,477
	Absent	2,220	12,733	14,953
Total		5,555	12,875	18,430

The deliverables for this project include:

- A list of possible erroneous (false positive) ingredient mappings (identified lexically);
- The list of all RxNorm clinical drugs with an asserted mapping to SNOMED CT, but no corresponding inferred mapping, along with an indication of the main type of error preventing the equivalence from being inferred.

Additionally, for each type of error, we provided the list of clinical drugs affected and suggestions for remediation.

- SPE:
  - A list of RxNorm clinical drugs with unmapped dose forms
  - A list of these dose forms
- MBO or MAI:
  - A list of RxNorm clinical drugs with missing basis of strength substance or active ingredient information
- ENM:
  - A list of RxNorm clinical drugs with erroneously unmapped ingredients,
  - A list of these ingredients (false negative mappings, with their mapping to a SNOMED CT substance we identified lexically).
- SSU:
  - A list of RxNorm clinical drugs with unmapped ingredients
  - A list of these unmapped ingredients (i.e., ingredients specific to RxNorm).
- ERR:
  - A list of RxNorm clinical drugs and their ambiguous dose forms
  - A list of these dose forms (with possible mappings to pharmaceutical dose forms in SNOMED CT)
- Semantic:
  - DFE: A list of RxNorm clinical drugs with the unmatched dose form.
  - BOS: A list of RxNorm clinical drugs with their unmatched BoSS. We also indicated whether or not the RxNorm ingredient and SNOMED CT substance are related by a modification relation.
  - SAI: A list of RxNorm clinical drugs with their actives ingredient(s). We also indicated whether or not the RxNorm ingredient and SNOMED CT substance are related by a modification relation.
  - SUM: A list of RxNorm clinical drugs with their unmatched unit(s) of measure. We also indicated whether the difference occurred in the denominator or numerator (or both), and whether it occurred in the concentration or presentation template (or both).
  - VAL: A list of RxNorm clinical drugs with their unmatched value(s). We also indicated whether the difference occurred in the denominator or numerator (or both), and whether it occurred in the concentration or presentation template (or both).

Figure 4 describes the distribution of errors identified through our analysis of the mappings.

Figure 4: Distribution of errors identified through our analysis of the mappings

4.3   Defining a course of action

Ingredient issues

The possible false positive and false negative mappings we identified lexically between ingredients and substances should be reviewed by RxNorm editors (ENM).

RxNorm ingredients with no mapping to SNOMED CT should be reviewed by SNOMED CT editors for inclusion into SNOMED CT (SSU).

Missing information about active ingredient or basis of strength substance in RxNorm clinical drugs should be added by RxNorm editors (MBO, MAI).

Clinical drugs for which there is a difference in basis of strength substance or active ingredient between RxNorm and SNOMED CT should be reviewed for accuracy by both the RxNorm and SNOMED CT editors (BOS, SAI).

Dose form issues

RxNorm dose forms specific to RxNorm should be mapped to SNOMED CT pharmaceutical dose forms by RxNorm editors (SPE)

Ambiguous RxNorm dose forms can be mapped to multiple SNOMED CT pharmaceutical dose forms (i.e., multiple definitions for the RxNorm clinical drugs) to maximize the chances if inferring equivalences between clinical drugs in SNOMED CT. However, RxNorm could consider refining the definition of the corresponding clinical drugs to precisely map to SNOMED CT (ERR).

Clinical drugs for which there is a difference in dose form between RxNorm and SNOMED CT should be reviewed for accuracy by both the RxNorm and SNOMED CT editors (DFE).

Strength issues

Clinical drugs for which there is a difference in unit or value between RxNorm and SNOMED CT should be reviewed for accuracy by both the RxNorm and SNOMED CT editors (SUM, VAL).

Updated integration result

Table 2 describes the distribution of the RxNorm clinical drugs according to their asserted and inferred mappings to SNOMED CT clinical drugs. In this table, the inferred mappings are obtained after curation of the mapping of ingredients and dose forms. The curation process resulted in inferring 174 additional mappings between clinical drugs.

Table 2: Distribution of RxNorm clinical drugs according to the source of mapping after the after curation of the mapping of ingredients and dose forms

		Asserted mappings		total
		Present	Absent	
Inferred mappings	Present	3,509	149	3,658
	Absent	2,046	12,726	14,772
Total		5,555	12,875	18,430

5 Acknowledgments