OCRx: Canadian Drug Ontology

Jean Noel NIKIEMA [a,b,1] , Man Qing LIANG [a,b] , Philippe DESPRÉS [c] and
Aude MOTULSKY [a,b]

[a] *Research Center, Centre hospitalier de l'Université de Montréal, Montréal, Canada*
[b] *Department of Management, Evaluation and Health Policy, School of Public Health,*
Université de Montréal, Québec, Canada
[c] *Département de physique, de génie physique et d'optique,*
Université Laval, Québec, Canada

Abstract. This paper describes the development and evaluation of a Canadian drug ontology (OCRx), built to provide a normalized and standardized description of drugs that are authorized to be marketed in Canada. OCRx aims to improve the usability and interoperability of drugs terminologies for a non-ambiguous access to drugs information that is available in electronic health record systems. We present the first release of OCRx that is described in Web Ontology Language and aligned to the Identification of Medicinal Product (IDMP) standards. For comparison purposes, OCRx is mapped to RxNorm, its US variant.

Keywords. Ontology, medication management, interoperability, IDMP

1. Introduction

Capturing real life information about how medication are prescribed, dispensed and administered is important, at both the individual (e.g., safety and efficacy) and jurisdictional levels (e.g. pharmacovigilance) [1]. With the computerization of health information systems, drug-related information needs to be produced, stored, and shared rapidly across various providers and organizations in a non-ambiguous way. As a result, the backbone for medication management is the standardized drug terminologies that are sustaining the interoperability of medication information between health information systems. Around the world, countless institutions, electronic health record (EHR) systems' builders and knowledge resources' providers have constructed or used diverse drug terminologies to meet their needs. To ensure the interoperability between the various tools in a jurisdiction, many countries have developed normalized descriptions of drugs. This is the case of RxNorm in the USA [2], and of Normalized Chinese Clinical Drug (NCCD) in China [3].

Indeed, the characteristics of drugs are often dependent on the jurisdiction in which they are marketed (non-commercial names: paracetamol vs acetaminophen, manufactured dose forms, approved therapeutics indications, strengths, etc.) [1,4]. Thus, each country must lead the standardization of its drug-related information, but also conceive these terminologies in a standardized way that can ensure interoperability across jurisdictions and countries. Thereby, in addition to the rules that the

---

[1] Corresponding Author: Jean Noel NIKIEMA, Research Center, Centre hospitalier de l'Université de Montréal; E-mail: jean.nikiema@umontreal.ca

conceptualization process must follow in order to ensure the quality of drugs terminologies' content [5], the IDMP (Identification of Medicinal Products) is nowadays the reference set of standards that drug terminologies must conform to [4]. In Canada, the terminologies used in EHRs and pharmacy systems are mainly provided by private companies. Additionally, the publicly available terminologies do not ensure a non-ambiguous and normalized way to access drug descriptions. As pointed out in [6], the characteristics of these drugs terminologies make the development of a standard terminology for medications a high priority in Canada. To meet that need, we built a drug ontology, OCRx (Canadian drug ontology), to ensure a normalized, standardized, and formal description of Canadian drugs. By its compliance with international standards, like IDMP, this effort has been made to ensure the interoperability but also the qualities of nationwide terminologies in Canada. This paper first describes the available public terminologies used to engineer OCRx, then presents the model and standardization process applied, and finally discusses our main findings.

2. Materials

To build OCRx, we used two public databases: Health Canada's Drug Product Database (DPD) and the Canadian Clinical Drug Dataset (CCDD). The DPD is maintained by Health Canada and contains a description of clinical drugs authorized to be marketed. Health Canada uses a Drug Identification Number (DIN), an eight-digit number, for every medication by a relevant strategy consisting of generating DIN code on the basis of 1) the active ingredient, 2) the manufacturer, 3) the strength, 4) the pharmaceutical form, and 5) the route of administration. However, only the lexical aspects of these definitional elements (i.e., characteristics that are used to define a concept) are used in this strategy, while the labels are mainly proposed by the manufacturer (e.g., non-commercial name, non-conventional form, etc.). Thus, a simple administrative change from the manufacturer (e.g., manufacturer name update) can lead to the creation of a new DIN. Using the framework in [5], it can be affirmed that, by design, DPD has a concept orientation, consistency and non-redundancy issues. Despite these limitations, DPD is the primary public source of drug information in Canada. Its content can be accessed through an API [7]. Because DPD does not contain a clinical drug concept, the CCDD was built by Canada Health Infoway to provide identifiers for human clinical drugs and to support the electronic prescription process. However, by building this dataset based solely on the lexical structure available in the DPD, Canada Infoway has been confronted with the same issues. The CCDD dataset is available online and can be downloaded in French and English [8].

3. Methods and Results

To construct OCRx, three main steps were applied: the definition of a model, the standardization of definitional elements, and the creation of axioms and an OWL (Web Ontology Language) file. Finally, the content of OCRx was compared with RxNorm as the main drug related terminology whose model is reused by many other terminologies [9].

3.1. Defining a Model

The OCRx model is based on four main concepts (Figure 1). The **Canadian clinical drug** is a clinical drug defined by its components and a pharmaceutical form. Each Canadian clinical drug is annotated by its marketing authorization number (DIN), defined as the **prescriptible clinical drug** (e.g., designating prescription and over-the-counter medications). The **component** is an active ingredient (substance) defined by its moiety form (substance), its strength, and the reference for the strength value (substance). The **form** is a basic form (e.g., tablet) with the precision of its route of administration (e.g., oral) and distinction (e.g., extended release).
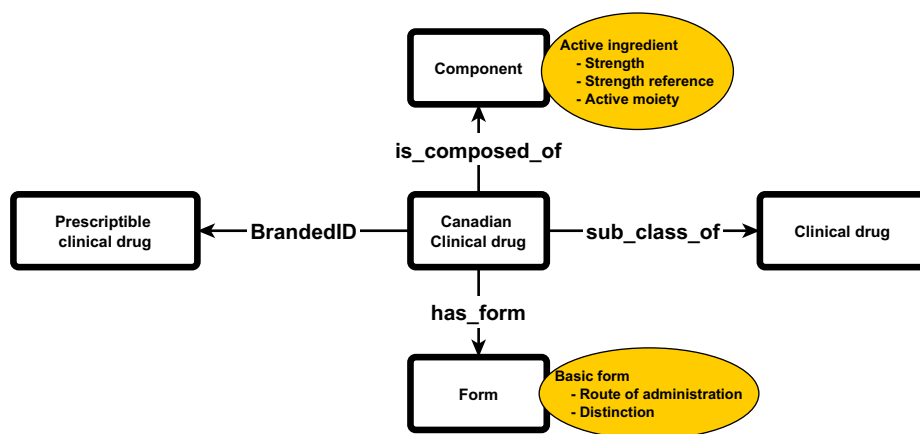

Figure 1. The OCRx model: A Canadian clinical drug is marketed using a specific ID and defined by its component (an active ingredient in a specific dosage) and form

3.2. Standardizing the Definitional Elements

We standardized each primary element used to describe clinical drugs.

1. **Substances**: A set of substances labels are acquired through the DPD API and CCDD release. We used the Global Ingredient Archival System (GINAS) developed by the US National Center for Advancing Translational Sciences [10] to identify variant in substance names and standardized those contained in the DPD and the CCDD. We then used the Unified medical language system (UMLS) to obtain different French synonyms. A unique identifier is attributed to labels declared similar through GINAS and the UMLS. Thus, 5,254 substances with 38,169 labels in French and English were retrieved from this process.
2. **Strength:** A home-made algorithm was created to standardize the strengths in the DPD (e.g., all the masses are converted in milligrams and all the volume in milliliters). For strength labels identified as synonyms by the algorithm (e.g., "0.015 MG", ".015 MG", "15 MCG", "15.0 MCG"), a unique identifier was created. OCRx contains 10,132 strength concepts for 10,882 labels.
3. **Form attributes:** Three annotators (MD, PharmDs) performed a manual standardization of distinction, route of administration and basic form and then created a unique identifier for these definitional elements. This standardization

provided 906 pharmaceutics forms including 150 basic forms, 95 routes of administration, 50 distinctions with labels in both English and French.

3.3. Creating an Ontology Pattern and Release

For each DIN, an identifier of "Canadian clinical drug" was generated based on the unique set of standardized definitional elements. Nine properties were created to ensure the relevant links between concepts and their definitional features. For example, **"is a modified form of"** is a transitive property used to link substances (e.g., link between base and salt), **"is composed of"** is a transitive property defined as a **"has active ingredient"** property and used to link "Canadian clinical drug" and "Component". Finally, the chain property **"is composed of"O"is a modified form of"** is a sub-property of "is composed of". "Canadian clinical drug" and "Forms" are represented as fully defined concepts. The others entities in OCRx are represented as primitive concepts. The formal structure of OCRx (ALCR DL expressivity) can be classified by Fact++ in 372.831 ms and Hermit in 83.812 ms. OCRx content can be browsed at ***http://ocrx.ca/***.

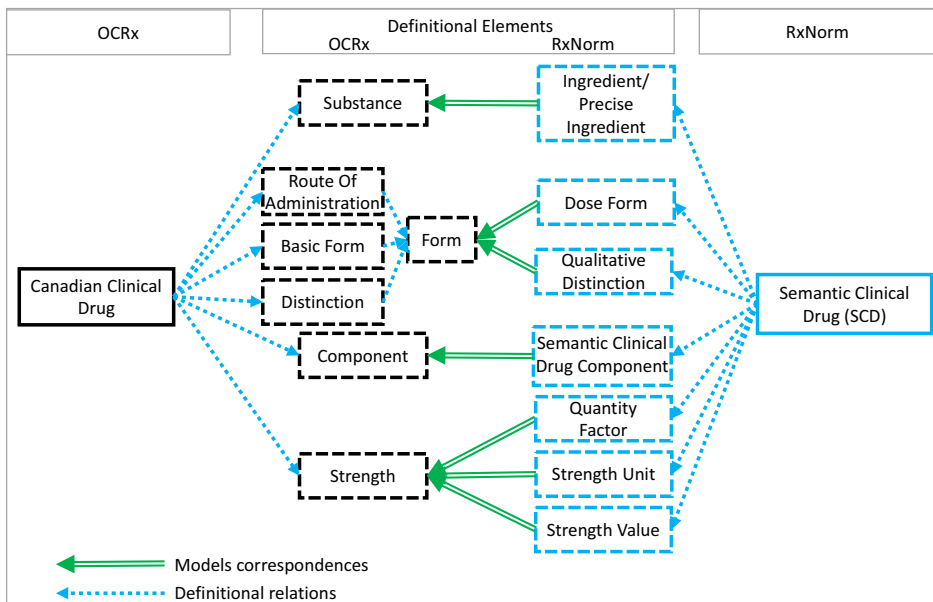## 3.4. Mapping to external resources

Figure 2. The OCRx model: A Canadian clinical drug is marketed using a specific ID and defined by its component (an active ingredient in a specific dosage) and form

The OCRx model was compared to the subset model of RxNorm describing nonproprietary names. We first look for equivalence between the models' elements (Figure 2). We then create lexical mappings of definitional features, and using DLqueries on the structures of OCRx, we retrieved the clinical drugs that are common. Thus, 3,386 substances are mapped to 3,518 Ingredients in RxNorm, 3,305 components are mapped to 2,974 clinical drug components, 129 forms are mapped to 69 Dose Forms, and 1,895 Canadian clinical drugs are mapped to 1,741 Clinical drugs in RxNorm.

4. Discussion-Conclusion

To our knowledge, OCRx is the first multilingual drug ontology conceived in DL language allowing a description without ambiguity of drug characteristics. The structure performs correctly and can be efficiently classified by usual reasoners. In the model, if "Prescriptible Clinical drugs" were represented as instances of "Canadian clinical drugs", the resulting structure would induce more than two hours of classification. Although this instance-description for DIN could be conceptually more relevant, it hinders the usability and scaling of the ontology because of the considerably greater time required to classify the structure. Therefore, using an annotation relation between "Prescriptible Clinical drugs" and "Canadian clinical drugs" to build this ontology represented a practical choice.

Like SNOMED CT [4] and RxNorm [9], OCRx is compliant with IDMP requirement and can easily be mapped to RxNorm concepts. OCRx resolves the intrinsic issues of existing drugs terminologies in Canada and allows multiple usages, from data integration, natural language processing, to drugs description in EHRs.

Despite the clear advantage of its structure, OCRx compliance to IDMP is limited for the strength description. Currently, the standardization process for strength description is made by a home-made algorithm. To improve its compliance with IDMP, the next update of the model will consist of introducing units of measures ontologies (e.g., the Unified Code for Units of Measure UCUM) to better represent the strength in components. The next steps for OCRx will also consist in integrating therapeutics indications, adverse reactions and possible drug interactions.

References

[1]   Motulsky A. Big Data Challenges from a Pharmacy Perspective. In: Big Data, Big Challenges: A Healthcare Perspective. Springer; 2019. p. 33–44.
[2]   Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm. Yearbook of medical informatics. 2018;27(1):129.
[3]   Wang L, Zhang Y, Jiang M, Wang J, Dong J, Liu Y, Tao C, Jiang G, Zhou Y, Xu H. Toward a normalized clinical drug knowledge base in China-applying the RxNorm model to Chinese clinical drugs. J Am Med Inform Assoc. 2018 Jul 1;25(7):809-818.
[4]   Bodenreider O, James J. The New SNOMED CT International Medicinal Product Model. In: ICBO; 2018.
[5]   Zhu X, Fan JW, Baorto DM, Weng C, Cimino JJ. A review of auditing methods applied to the content of controlled biomedical terminologies. Journal of Biomedical Informatics. 2009;42(3):413–425.
[6]   Motulsky A, Sicotte C, Gagnon MP, Payne-Gagnon J, Langué-Dubé JA, Rochefort CM, Tamblyn R. Challenges to the implementation of a nationwide electronic prescribing network in primary care: a qualitative study of users' perceptions. J Am Med Inform Assoc. 2015 Jul;22(4):838-48.
[7]   Drug Product Database (DPD) API Guide, Available at: https://health-products.canada.ca/api/documentation/dpd-documentation-en.html
[8]   Canadian Clinical Drug Data Set, https://infocentral.infoway-inforoute.ca/en/standards/canadian/ccdd
[9]   Nikiema JN, Bodenreider O. Comparing the representation of medicinal products in RxNorm and SNOMED CT–Consequences on interoperability. In: ICBO; 2019.
[10] The Ginas Project, https://gsrs.ncats.nih.gov/#/