Towards a foundational representation of potential drug-drug interaction knowledge

Mathias Brochhausen\textsuperscript{1}, Jodi Schneider\textsuperscript{2}, Daniel Malone\textsuperscript{3}, Philip E. Empey\textsuperscript{4}, William R. Hogan\textsuperscript{5} and Richard D. Boyce\textsuperscript{6}

\textsuperscript{1} Division of Biomedical Informatics, University of Arkansas for Medical Sciences, 4301 W. Markham St., #782, Little Rock, AR, 72205-7199
\textsuperscript{2} Web-instrumented man-machine interactions, communities and semantics group, INRIA Sophia Antipolis – Méditerranée, France
\textsuperscript{3} College of Pharmacy, University of Arizona, Tucson, AZ, USA
\textsuperscript{4} Department of Pharmacy and Therapeutics, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA
\textsuperscript{5} Department of Health Outcomes and Policy, College of Medicine, University of Florida, Gainesville, FL, USA
\textsuperscript{6} Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA

mbrochhausen@uams.edu, jschneider@pobox.com, malone@pharmacy.arizona.edu, pempey@pitt.edu, hoganwr@ufl.edu, rdb20@pitt.edu

Abstract. Inadequate representation of evidence and knowledge about potential drug-drug interactions is a major factor underlying disagreements among sources of drug information that are used by clinicians. In this paper we describe the initial steps toward developing a foundational domain representation that allows tracing the evidence underlying potential drug-drug interaction knowledge. The new representation includes biological and biomedical entities represented in existing ontologies and terminologies to foster integration of data from relevant fields such as physiology, anatomy, and laboratory sciences.

Keywords: Potential drug-drug interactions, ontologies, knowledge management

1 Introduction

Every year, many thousands of people are harmed by exposure to two or more drugs for which there exists a known interaction potential. Exposure to such “potential drug-drug interactions” (PDDIs), are a significant source of preventable drug-related harm, leading to clinically important events in 5.3\% - 14.3\% of inpatients, and accounting for 0.02\% to 0.17\% of the 129 million emergency department visits that occur in the U.S. each year\cite{1}\cite{2}. Multiple defenses exist in the healthcare system to prevent patient harm from PDDIs including clinician knowledge, computer screening,
and monitoring. Each defense depends on complete, accurate, and current knowledge of what drugs have the potential to interact, and the most appropriate methods for managing patients when exposure to a PDDI is unavoidable [3]. However, most sources of clinically-oriented PDDI knowledge disagree substantially in their content, including about which drug combinations should never be never co-administered. For example, only one quarter of 59 contraindicated drug pairs were listed in three PDDI information sources[4], only 18 (28%) of 64 pharmacy information and clinical decisions support systems correctly identified 13 PDDIs considered clinically significant by a team of drug interaction experts[5], and four clinically oriented drug information compendia agreed on only 2.2% of 406 PDDIs considered to be “major” by at least one source[6].

A key factor underlying the existing disagreements among sources of drug information that are used by clinicians is the inadequate representation of PDDI evidence and knowledge. In practice, organizations that provide PDDI information as part of their information services employ an expert or panels of experts (editorial boards) to search, evaluate, synthesize, and stay current with evidence. The process involves applying some criteria to judge whether a drug combination could lead to an interaction, what impact it might have on exposed patients, and how to best manage patient exposure. In the current paradigm, these individuals or groups must search across multiple information sources, including the scientific literature, drug product labeling, and documents submitted to regulatory groups during the drug development/approval process. There is significant variation across drug knowledge bases with respect to ratings of specific drug pairs and currency. Moreover, the available sources rarely include first-hand clinical experience, information that can help contextualize management recommendations[7]. In addition, those multiple sources are currently not created in a way that fosters semantic integration of their data at a later stage. This leads to inefficient and discordant approaches to the acquisition of PDDI evidence and synthesis of that evidence into knowledge. The result is that there is general disagreement among drug information systems about what PDDI exist and their clinical importance.

A goal of the “Addressing gaps in clinically useful evidence on drug-drug interactions” project is to identify the core components of a new PDDI knowledge representation paradigm that addresses these issues. As we describe below, the project makes a fundamental distinction between assertions of PDDI knowledge and the evidence that supports or refutes such assertions. The central thesis of the project is that a framework for representing PDDI assertions and evidence as interoperable Linked Data[8] will enable a more integrated approach to the acquisition and synthesis of PDDI evidence into knowledge. Linked Data methodologies should be used to semantically integrate the various relevant sources of PDDI evidence so that experts can more easily retrieve all relevant evidence items. This will lead to more complete, accurate, and current PDDI information provision to any single evidence board than is possible with current resources.

The proposed framework requires a new foundational representation of PDDIs that covers the material entities and processes in the domain of discourse for PDDI evidence and knowledge claims. The representation will enable the integration of drug
interaction mechanisms, effects, risk factors, severity, and management options with the chemical and pharmacological properties (e.g., chemical structure, function, pharmacokinetic and pharmacodynamic properties) of the interacting drugs. This paper specifies the design requirements for such a foundational representation that we are calling the Drug-drug Interaction and Drug-drug Interaction Evidence Ontology (DIDEO). Section 2 provides clinical background. Section 3 discusses the basic design principles and decisions for the new ontology. Finally in section 4, we show that the classes in DIDEO are sufficient to represent a concrete example of PDDI evidence selected from the Drug Interaction Knowledge Base.

2 Background

2.1 A conceptual framework for clinically useful PDDI Knowledge and Evidence

There is a rather complex relationship between the evidence that establishes a PDDI, and information that can help clinicians accurately assess the risk of exposure within a given patient[7]. The foundational model we envision would benefit from an explicit conceptual model of that relationship. Eric van Roon et al. proposed a conceptual model of PDDI information using the definition that clinically-useful PDDI information is that which helps discern whether some action should be taken with respect to a PDDI (Figure 1)[9]. Evidence for, or against, the existence of a PDDI is an important component in that model, along with consideration of patient risk factors, the potential severity of an adverse event that could be caused by exposure, and prior experience with exposure in relevant patient populations. While the van Roon model is not considered a standard for representing PDDI knowledge, it captures the essence of recommendations by other PDDI experts[10][11], including developers of PDDI databases in the United States and Europe[12][13].

The van Roon model helps to conceptually outline the principal information domains for clinically-useful PDDI knowledge. We think that it is also important to consider the relationship between PDDI evidence and claims of PDDI knowledge established by evidence. The evidence for, or against, PDDI assertions is dynamic and of varying robustness to various forms of bias. For example, in prior work on the Drug Interaction

![Fig. 1. The four types of information used by van Roon et al. to determine if a PDDI warrants clinical action. AE – Adverse Event](image1)

![Fig. 2. Distinctions between evidence, knowledge, and reasoning with respect to PDDIs](image2)
Knowledge Base (DIKB) [14][15][16][17], the editorial board considered certain pharmacologic assertions written in a FDA guidance to industry[18][19][20] useful as PDDI evidence. The assertions reflected the state of science at the time the documents were published. The guidance has been updated, from 1999 to 2006 to 2012, each update leading to changes in the DIKB evidence base.

Based on these observations, it’s possible to conceive of PDDI evidence board as sociotechnical reasoning system that manages both an evidence base and a knowledge base (Figure 2)[21]. This implies that the foundational PDDI knowledge representation we envision would find application within three specific contexts:

1) within the evidence base, by defining the four types of information from van Roon’s model (Figure 1): evidence that can be used to establish the existence of a PDDI, patient risk factors, the potential severity of an adverse event that could be caused by exposure, and prior experience with exposure in relevant patient populations;

2) within the knowledge base, by representing the entities explicit within PDDI assertions (e.g., “drug X interacts with drug Y”) and pharmacologic assertions that can be used to infer PDDIs (e.g., “drug X inhibits enzyme Q which is important for the clearance of drug Y from the body”); and

3) within the reasoning system, by constraining the inference activities of the evidence board so that inferred knowledge is logically consistent with all of the other assertions in the knowledge base.

These distinctions can be illustrated using the artifacts used to support the DIKB. Underlying the DIKB’s evidence base are specific examples of evidence types; these evidence types are outlined in a draft online document[22]. Meanwhile the DIKB’s knowledge base contains PDDI and pharmacologic assertions; these assertions address competency questions that were identified during prior work on the DIKB, which can be found in a different draft online document[23].

2.2 Related work

Currently, there are two ontologies built specifically for the domain of drug-drug interactions: the Drug Interaction Ontology (DIO)[24] and the Drug-drug Interaction Ontology (DINTO)[25]. Both provide insights that are valuable for representing the domain. However, neither was designed with the perspective outlined above. Nor do they allow for a consistent and scalable representation of the ontological distinctions relevant to representing clinically useful drug-drug interaction assertions, the drugs involved, and the supporting or refuting evidence. We introduce these ontological distinctions in the course of discussing the existing ontologies.

The DIO is an ontology of drug interactions developed with the goal of predicting drug interactions[24]. While DIO[24] is inspired by both Basic Formal Ontology (BFO) and the NCI Thesaurus (via UMLS), it is not aligned with either one. Although the DIO [24] specifically refers to BFO’s distinction between continuants and processes (occurrents)[26], the BFO’s representation of process (its definition and entity URI), is not reused in the OWL implementation of DIO accompanying the aforementioned paper[27]. Rather, within the DIO OWL file, process is defined as: “A se-
quence of events which produces some outcome” [27], reusing the CUI and definition from the National Cancer Institute Thesaurus (NCIT)[28]. But the DIO OWL file includes several axioms that are inconsistent with both the NCIT definition of process and the BFO’s representation of process. In particular, DIO specifies necessary conditions for processes, such as

- hasEnableTriggerParticipant min 1 Thing
- hasResultantPopulationChange min 1 Thing
- hasResultantPopulationChange only Increased

A possible explanation for this inconsistency is that the developers of DIO intended to map their domain representation to those found in the UMLS (NCIT) and other terminologies and ontologies rather than actually reuse them. However, the reuse of terms should always be accompanied by ensuring that the intended meaning and the ontological commitments of source and target resource match. (We will elaborate on the role of ontological commitments in Section 3.)

Another shortcoming of DIO is that it does not represent roles. Each instance of a chemical is a drug, regardless of whether its dosage or formulation allows it to act as a drug. We find, for instance, that:

\[
\text{Capecitabine}
\]
\[
\text{rdfs:subClassOf Drugs}
\]
\[
\text{rdfs:subClassOf DrugOrMetabolite}
\]
\[
\text{rdfs:subClassOf Chemicals}
\]

But active ingredients can only bear a role as a drug in a specific dose and in conjunction with excipients[10]. However, it is by now standard accepted practice in numerous drug terminologies and ontologies to carefully distinguish among drug products, their ingredients, and the molecules that constitute those ingredients[29][30][31][32]. Most recently, Hogan et al. in this regard showed that assigning therapeutic properties to active ingredients disregards the effect of dose form and therefore leads to mistakes that contradict scientific knowledge (e.g., oral vancomycin treating bacterial endocarditis)[30].

The second ontology that we took into consideration is the Drug Interaction Ontology (DINTO). DINTO is intended “to represent all possible mechanisms that can lead to a drug-drug interaction. The ontology provides the general pharmacological principles of the domain”[25]. The developers have provided a version of DINTO that is an extension of BFO[33]. The key limitation of DINTO with respect to our goals is that DINTO does not represent potential drug-drug interactions at all, but only drug-drug interactions (DDIs). Representing PDDIs, as we aim to do, is quite different from representing a DDI. For each individual instance of a drug-drug interaction it is possible to specify the individual patient who suffered from its effects. However, this is not possible for all instances of PDDIs, because some of them are not actualized. DIDEO will be based on a novel definition of PDDI (Section 3).

While DINTO does not represent PDDIs at all, the way it represents the actual occurrences of DDIs and information about those occurrences is problematic. DINTO specifies a subclass of DDIs named \textit{DDI described in a database}. The members of
this class are intended to be DDIs\(^1\) that are linked to a database by the *is described in* relation. According to the DINTO OWL file\([33]\) the class ‘DDI in database’ is intended to “represent those DDIs imported in DINTO from the DrugBank database\([34]\)\([35]\) with the purpose of distinguishing them from those inferred from the ontology”. Notably, the information loaded from DrugBank will not be about drug-drug interactions, but about PDDIs, as \([35]\) clearly indicates. This demonstrates that DINTO does not provide ways to distinguish information about actual DDIs from evidence pointing at drug co-medications that are suspected to lead to unwanted effects.

3 Methods

Building from the pioneering work of the DIO and DINTO we propose to develop a new ontology, DIDEO. The DIDEO will address the aforementioned limitations of the two ontologies while being in alignment with the van Roon conceptual model. Moreover, the ontology will comply with principles of good practice in ontology development, such as formulated by the OBO Foundry\([36]\), for instance:

*Reuse of pre-existing resources* – Because integration of data is among the key rationales for using ontologies, it is crucial to use Unique Resource Identifiers to refer to the same entities even across domains. We strive to reuse entities from pre-existing ontologies wherever reasonable. Reuse of entities may be limited by the fact that the basic ontological commitments of the source and the target ontology need to be the same. For instance, an ontology that defines ‘drug’ as ‘a chemical entity that bears a drug role that is realized by its use in a pharmacotherapy’ cannot import an individual drug, for instance ‘acetaminophen’ from an ontology that defines drugs as chemicals that are used in pharmacotherapy\(^2\). In our example both ontologies represent ‘drug’, but each representation comes with a different ontological commitment. One way to assure consistency of ontological commitment is to select entities from ontologies using the same upper ontology.

*Use of an upper ontology relevant in biomedical informatics* – The DIDEO should support the integration of drug-drug interaction data with data on other biomedically relevant phenomena, for example proteins, protein interactions, laboratory methods and clinical studies. We reuse entities from the Drug Ontology (DRON) \([29]\)\([30]\)\([37]\) the Ontology of Biomedical Investigation (OBI)\([38]\)\([39]\), the Gene Ontology (GO)\([40]\)\([41]\) and the Information Artifact Ontology (IAO)\([42]\) These ontologies are all listed on the OBO Foundry\([43]\) webpage. The reuse of DRON, OBI and IAO commits the DIDEO to use the Basic Formal Ontology (BFO) \([26]\)\([44]\) as the upper ontology. The Gene Ontology is not using any upper ontology, but multiple classes from GO have been subsumed under BFO classes in the aforementioned ontologies.

---

\(^1\) The axiomatization of the class actually falls short of specifying that, since being a DDI is not part of its necessary and sufficient condition.

\(^2\) The aspirin in my medicine cabinet is not a drug according to the second definition given above since it is not participating in any pharmacotherapy, yet it is most certainly a drug according to the first definition.
Representation of biological and biomedical entities was one of the use cases in the development of BFO[26]. In addition, BFO provides well-documented representations for roles, functions and dispositions[45], which are also relevant for biological and biomedical phenomena. Hence we plan to use BFO 2.0 as our upper level and we will import existing OWL entities for reuse using the MIREOT methodology[46] implemented in the MIREOT Protégé plugin[47]. One open question when using MIREOT to import terms from pre-existing ontologies is how to track changes in the source ontologies. MIREOT [46] relies on OBO Foundry’s internal (non-automatic) monitoring process, which might not be an optimal solution, especially when the methodology is applied outside the Foundry.

Compliancy with relevant standards of drug representation – The initial development of DIDEO re-uses the drug representation of the Drug Ontology[29] (DRON) which is based on RxNorm[32]. For a PDDI ontology, active ingredients must not be assigned the status of drugs, because the excipients, route of administration, and dose impact the potential for, likelihood of, and severity of interactions. DRON provides ontologically sound representations of Clinical Drugs and Branded Drugs. These representations contain information about dosage and intended route of administration. In addition, DRON provides information about drug ingredients that is linked to the Chemical Entities of Biological Interest ontology [29][30][48].

Community-driven development – Once the initial OWL version of DIDEO is created it will be made publicly available and the project will be continued as an open source project. In addition, we aim to build a community of consumers/contributors to help us create, expand and maintain the ontology. We have already reached out to the developers of DINTO. Since DINTO is using BFO as the upper ontology just as DIDEO we want to investigate possibilities to align our efforts.

4 Results

To appropriately represent PDDI knowledge and its evidence, a definition of PDDIs is crucial. We start our definition by making a basic ontological categorization based on BFO 2.0[49]. The most basic ontological categories of BFO are independent continuant (such as material entities), generically dependent continuants (such as information content entities), specifically dependent continuants (such as qualities and dispositions) and occurrents (such as processes)[26]. Consider an individual PDDI. We might be tempted to categorize it as an occurrent, since the term “interaction” points to a process. However, a PDDI is not an actual process. It is also not a potential or disposition inhering in a substance that may or may not be realized. Representing PDDI in that way, would neglect the fact, that in PDDI research we are collecting information about occurrences that could be drug-drug interactions. Rather, a PDDI is a piece of information about the possible effects of a certain event, for instance the co-administration of azithromycin and ergot alkaloids. We propose the following definition for PDDI:

"A potential drug-drug interaction (PDDI) is an information content entity that specifies the possibility of a drug-drug interaction based on either reasonable extrapolation about drug-drug interaction mechanisms or a data item
created by clinical studies, clinical observation or physiological experiments.”

From this starting point it is crucial to represent the informational bases of PDDIs, namely a) reasonable extrapolation, b) physiological observations from clinical studies and c) drug-drug interaction observational data and d) mechanistic assertions that are useful for inferring drug-drug interactions (derived either from clinical studies or from various experiments, such as inhibition and transport protein experiments)

The aim of DIDEO is to adequately represent all types of PDDI evidence, as well as their differing bases. Not all instances of PDDI evidence data are based on actually observing a drug-drug interaction. In many instances, it is unclear from the evidence whether any actual interaction between the object and precipitant drug has occurred. Hence, it would not be appropriate for us to code the ontology in a way that implies, from the existence of the PDDI evidence, the existence of at least one instance of the specific drug-drug interaction. These bases imply the existence of specific physiological processes, drugs, drug components, and in some cases even DDIs. The following example describes a case of PDDI without evidence for an instance of the actual DDI:

Assume the class ‘potential azithromycin-ergot alkaloid interaction evidence data’ exists in an OWL ontology. To link the data item to actual physiological processes we could axiomatize that each element of this class is about some element in the class ‘azithromycin-ergot alkaloid interaction’. But our axiom will then imply that at least one element of the latter class exists.

This existential import can be avoided by using a feature novel in OWL2 called ‘punning’. ‘Punning’ enables users to assign the same name to an OWL class and an OWL individual, allowing the use of the individual when referring to the type and the use of the class when referring to individuals or aggregate of individuals. Despite the two entities bearing the same name, no cross-inferences are made when reasoning[50]. Thus, punning would allow stating that a PDDI is about at least two types of drugs, without affirming that each individual entity is about one individual portion of that drug.

Figure 3 shows a DIDEO representation (without punning) of a clinical study potentially useful as evidence for a mechanistic assertion that could be used for inferring drug-drug interactions. The figure shows how information about the drug and enzyme involved in the study can be traced from the study data item. It would also be possible to track the type of clinical study. Some ontological commitments and design decision depicted in Figure 3 warrant more detail:

• A particular simvastatin metabolism process is the proper occurrent part of a drug metabolism assay. In natural language we might say that the metabolism participates in the assay. But one of the ontological commitments of BFO is that the ’participates in’-relation only holds between a continuant and an occurrent[51]. To simply say that the metabolism and the assay temporarily overlap would not be sufficient here. Many processes overlap in time, without being interrelated in any other way.
• When we represent substances that are referred to by mass nouns, we talk about a portion of that substance (e.g. the simvastatin metabolism has a portion of simvastatin as a participant). This is inline with the practice used in numerous OBO Foundry ontologies, to distinguish between a specific instance of a portion of the entity and the term denoting its type[52].
Both the portion of simvastatin and the portion of CYP3A4 are affirmed to participate in the process at some times. Claiming their participation at all times we would exclude the possibility that there is no CYP3A4 available to participate in the metabolism, while the final part of the process are still occurring. Moreover, ‘participate at some time’ is not a negation of ‘participate at all time’, but, entails the latter [53].

Assays of a simvastatin metabolism establish that CYP3A4 is the bearer of a disposition called ‘drug metabolism enabler disposition’ that enables the metabolism, and is realized by that metabolism.

The outcome of the assay is a simvastatin metabolism data item. Data items are defined as “an information content entity that is intended to be a truthful statement about something (modulo, e.g., measurement precision or other systematic errors) and is constructed/acquired by a method which reliably tends to produce (approximately) truthful statements”[42] . The simvastatin metabolism data item is a member of the class of data items that are specified output of some drug metabolism assays.

5 Conclusion

This paper offers justification for a new foundational domain representation for PDDI knowledge and describes the initial steps toward its development. This new semantic model, the Drug-drug Interaction and Drug-drug Interaction Evidence Ontology (DIDEO), is motivated by the needs of experts who must search, evaluate, and synthesize PDDI evidence into knowledge claims. The results reported in this paper form a foundation for the further development of DIDEO. We will now start to implement DIDEO in OWL in order to test its applicability with respect to competency questions [23] specified in as part of the “Addressing gaps in clinically useful evidence on drug-drug interactions” project. During implementation we will seek to coordinate our efforts with the developers of DINTO and DRON.

Acknowledgments: For all authors: This project is supported by a grant from the National Library of Medicine: “Addressing gaps in clinically useful evidence on drug-drug interactions” (R01LM011838-01) and the National Institute of Aging “Improving medication safety for nursing home residents prescribed psychotropic drugs” (K01 AG044433-01). The authors thank Michel Dumontier and Alan Ruttenberg for their valuable comments, which have significantly improved the paper.

For PE: This work is supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number KL2TR000146.

For DM: This work is partially supported by the Agency for Healthcare Research and Quality (AHRQ) Grant No. 1R13HS021826-01 (Malone DC-PI).

For JS: This work was carried out during the tenure of an ERCIM “Alain Bensoussan” Fellowship Programme. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no 246016.
Fig. 3: Representation of instances derived from DIKB in the new semantic model, DIDEO. (Boxes represent OWL named classes; Text in boxes gives the label for the class and for imported terms the class URI, diamonds represent individuals, arrows represent object properties or rdfs:subClassOf (i.e., the “is a” relation)).
References


33. The Drug-Drug Interaction Ontology, https://4625527d7e0f0589865f115fb0e87fe18bef216f.googledrive.com/host/0B-7Po9tR1KLUkNpNdFEmcG44RjA/DINTO_1BFO.owl. Last accessed: 10/01/2014