Acquiring and representing drug-drug interaction knowledge as claims and evidence

Jodi Schneider[1] & Richard D. Boyce[2]

1- School of Information Sciences, University of Illinois at Urbana-Champaign

2- Department of Biomedical Informatics, School of Medicine, University of Pittsburgh

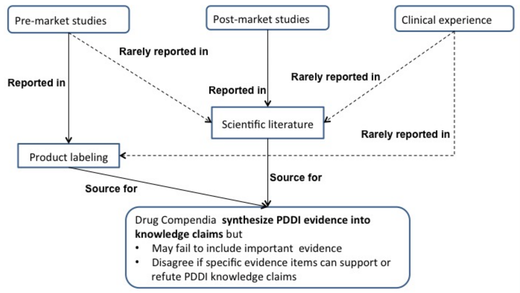
School of Information Sciences
The iSchool at Illinois

Problem

Potential drug-drug interactions are a significant source of preventable drug-related harm. The drug information sources clinicians use are disconcordant: **Most drug information sources disagree substantially in their content** (e.g. Abarca et al. 2003, Wang et al. 2010, Saverno et al. 2011)**. This problem has persisted for more than a decade** (e.g. Ayvaz et al. 2015, Ekstein et al. 2015) despite extensive editorial work on the part of each drug information source. This is in part because:

(1) There is no standard, agreed upon method for assessing evidence about drug-drug interactions.

(2) Knowledge claims and evidence about drug-drug interactions are distributed across multiple sources: pre-market studies, post-market studies, and clinical experience.



Aim

In this work, we address the distributed nature of drug-drug interaction knowledge, by developing a computable representation for claims and evidence about drug-drug interactions. Our goal is to support:

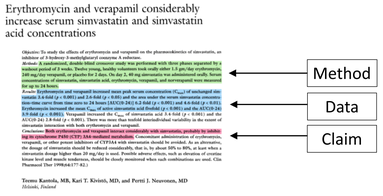
(a) Knowledge acquisition from full-text natural language

(b) Search and retrieval of all evidence.

We are applying this representation to acquire claims and evidence about pharmacokinetic interactions for 65 drugs. This will help us design a search portal, to test whether computable representations of knowledge claims and evidence can improve search and retrieval of potential drug-drug interactions.

Longer term, we will test whether this can help reduce the disconcordance between different drug information sources.

We model knowledge as claims supported by evidence.

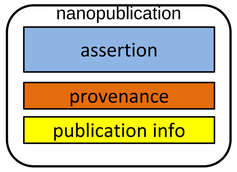

Erythromycin and verapamil considerably increase serum simvastatin and simvastatin acid concentrations

Method
Data
Claim

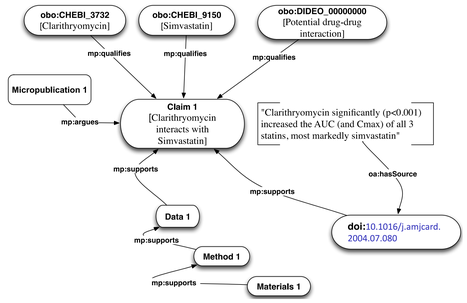
Approach

Designing a data model for representing claims and evidence

1. **Identify key ontologies relevant for claims and evidence.**
   - Nanopublications Ontology represents settled science: Each formalized claim (assertion) is wrapped in provenance and publication info.

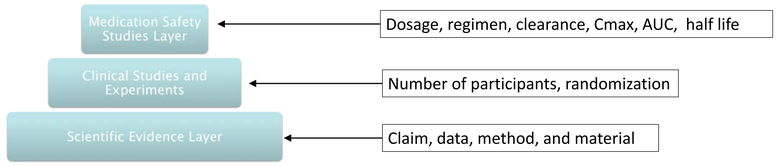


   - Micropublications Ontology represents claims and evidence. It views a scientific paper as a network of claims supported by data, methods and materials. Claims, data, methods, and materials can be text, images, or multimedia: anything the Open Annotation Ontology can reference.



2. **Identify key domain ontologies to reuse.**
   - Ontology of Biomedical Investigations
   - Chemical Entities of Biological Interest
   - Drug Ontology

3. **Conceptualize 3 layers and determine what belongs in each layer:**



4. **Formalize key terms about drug-drug interactions in a new ontology, DIDEO.**
For instance, "potential drug-drug interaction" gets obo:DIDEO_00000000

DIDEO: formalizing medication safety studies

As an OWL ontology, DIDEO supports querying as well as reasoning. For instance, curators will only need to enter a few facts in order for the scientific method of a study to be automatically determined.

DIDEO uses ontological realism to distinguish:
- Potential vs. actual drug-drug interactions.
- Inferred vs. observed interactions.

5. **Create micropublications claims and nanopublication assertions.**
Micropublication M: "Clarithromycin interacts with simvastatin"
Nanopublication assertion N: obo:CHEBI_3732 obo:DIDEO_00000000 obo: CHEBI_9150

6. **Connect natural language quotations and their information-retrieval friendly versions.**
M mp:formalizedBy N .
N mp:formalizes M .

Acquiring claims and evidence

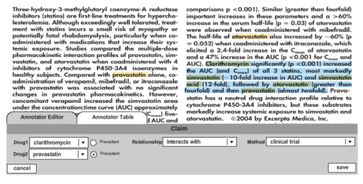
1. Formulate claims of interest.
"Clarithromycin interacts with simvastatin".

2. Identify relevant source documents.
Source documents include FDA-approved drug product labels and full-text research papers (clinical trials and case reports).

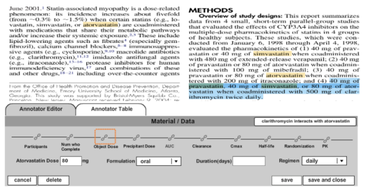
3. Experts assess quality & relevance of source documents.
Experts check that documents meet inclusion criteria. Experts find relevant claims, methods, and results.

4. Pre-annotation by computer text mining.
Source documents are pre-processed to find drug mentions, using named entity recognition algorithms.

5. Human curators annotate full-text documents.
(a) The curator highlights the claim.
(b) The curator enters the claim and scientific method.



(c) The curator is prompted to add data based on the method.



Annotation Results

From prior work, we have transformed
- 410 assertions
- 519 evidence items

Annotators have also identified evidence in
- 158 non-regulatory documents (including full-text research articles)
- 27 FDA-approved drug labels

This leads to an additional:
- 230 assertions of drug-drug interactions in non-regulatory documents
- 609 evidence items relating to potential pharmacokinetic drug-drug interactions from 27 FDA-approved drug labels

Future Work

- Build an information portal that supports clinical pharmacists and drug information professionals in retrieving the claims and evidence.
- Test the information portal in a task-based, within-subject, user study. Measure the completeness of the information experts retrieve with our information portal compared to current state-of-the-art retrieval tools.
- Test the feasibility of authors annotating their own claims and evidence.
- Enable annotation beyond PubMed Central open access HTML.
- Use rulesets of "belief criteria" to transform evidence to a knowledge base.

Ontologies, Data, and Websites

DIDEO: The Potential Drug-drug Interaction and Potential Drug-drug Interaction Evidence Ontology https://github.com/DIDEO/DIDEO

Contributions to Micropublications Ontology (formalizes/formalizedBy): https://github.com/dbmi-pitt/DIKB-Micropublication/blob/master/data/mp_1.18.owl

Drug Interaction Knowledge Base website & discussion forums http://dikb.org and http://forums.dikb.org

Publications and Presentations

1. Jodi Schneider, Mathias Brochhausen, Samuel Rosko, Paolo Ciccarese, William R. Hogan, Daniel Malone, Yifan Ning, Tim Clark and Richard D. Boyce. "Formalizing knowledge and evidence about potential drug-drug interactions." *International Workshop on Biomedical Data Mining, Modeling, and Semantic Integration* at *International Semantic Web Conference 2015* http://ceur-ws.org/Vol-1428/BDM2I_2015_paper_10.pdf

2. Jodi Schneider, Paolo Ciccarese, Tim Clark and Richard D. Boyce. "Using the Micropublications ontology and the Open Annotation Data Model to represent evidence within a drug-drug interaction knowledge base." *4th Workshop on Linked Science* at *International Semantic Web Conference 2014* http://ceur-ws.org/Vol-1282/lisc2014_submission_8.pdf

3. Mathias Brochhausen, Jodi Schneider, Daniel Malone, Philip E. Empey, William R. Hogan and Richard D. Boyce "Towards a foundational representation of potential drug-drug interaction knowledge." *First International Workshop on Drug Interaction Knowledge Representation* at the *International Conference on Biomedical Ontologies 2014* http://ceur-ws.org/Vol-1309/paper2.pdf

4. Mathias Brochhausen, Philip E. Empey, Jodi Schneider, William R. Hogan, and Richard D. Boyce. Adding evidence type representation to DIDEO. ICBO 2016 http://jodischneider.com/pubs/icbo2016.pdf

5. Jodi Schneider, Samuel Rosko, Yifan Ning, and Richard D. Boyce. "Towards structured publishing of potential drug-drug interaction knowledge and evidence. Poster presentation at: the Pittsburgh Biomedical Informatics Training Program 2015 Retreat. Pittsburgh, PA, August 20, 2015. http://dx.doi.org/10.6084/m9.figshare.1514991

6. Jodi Schneider and Richard D. Boyce "Medication safety as a use case for argumentation mining". Dagstuhl Seminar 16161: Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments, Dagstuhl, Germany, April 19, 2016 http://www.slideshare.net/jodischneider/medication-safety-as-a-use-case-for-argumentation-mining-dagstuhl-seminar-16161-2016-0419

FUNDED BY: training grant T15LM007059 from the National Library of Medicine/National Institute of Dental and Craniofacial Research and by R01LM011838 from the National Library of Medicine. ORIGINALLY PRESENTED AT: NLM Informatics Training Conference, Columbus, Ohio, June 2016