Semantic Web Scientific Discourse and Natural Language Processing to Enhance Drug-mechanism Knowledge in Drug Product Labels

A W3C HCLS IG Use-Case proposal
The problem – Incomplete or missing information in drug package inserts

• Example:
  – “Both clopidogrel and ticlopidine significantly inhibited the CYP2B6-catalyzed bupropion hydroxylation. Patients receiving either clopidogrel or ticlopidine are likely to require dose adjustments when treated with drugs primarily metabolized by CYP2B6.” [1]

• Search bupropion drug package inserts for “clopidogrel”
  – No mention at all in Aplenzin ER insert [2]
  – Mention in the generic tablet insert [3], but refers only to hypothetical interaction

Other examples of missing information

• Cimetidine inhibition of CYP3A4 and CYP1A2 metabolism

• Fluconazole inhibition of CYP3A4 metabolism
  – Noted in an FDA draft guidance [1] but not directly in the fluconazole solution PI [3]
    • Inferable from the stated interaction with midazolam

• See others like this in [4]

Other examples cont.

• The PI for citalopram [1] notes age-related pharmacokinetic changes
  
  “...subjects ≥ 60 years of age were compared to younger subjects in two normal volunteer studies. In a single-dose study, citalopram AUC and half-life were increased in the elderly subjects by 30% and 50%, respectively”

• However, clearance (Cl) would be the preferred measure of age-related change – is there literature on this?
  
  – Yes! [2-4]

So what?

• Aren’t package inserts for regulatory purposes only?
  – No – they are primarily intended to be a reference for prescribing clinicians [1]
• Who reads the package insert anyway?
  – 14% of prescribers that were surveyed reported using the package insert for information on drug-drug interactions (DDIs) [2]
  – Also, we have found that the agreement among drug information databases on DDIs is strongly correlated with mention in the package insert

Is there any requirement that the inserts be kept up to date?

• Kind of...
  – 2006 regulations explicitly require that studies of drug-drug interactions, metabolic pathways, and special populations be discussed [1]
  – Also, CFR 21 201.56 (a)(2) states "the labeling must be updated when new information becomes available that causes the labeling to become inaccurate, false, or misleading“ [2]
  – But, one could argue that incomplete information is not the same as inaccurate information
  – Very little interest from pharma in maintaining generic package inserts

What are we suggesting then?

- Providing a “mash-up” view of package inserts that alerts the reader to up-to-date information would be valuable:
  - Instant access to more complete evidence
  - Quick identification of gaps in knowledge on a drug
- For example:

http://dbmi-icode-01.dbmi.pitt.edu/pi-mashup-example/dikb-label-mashup-escitalopram.html
This could be accomplished with Semantic Web technology

- We developed a very simple prototype [1]
  - Used information in the Drug Interaction Knowledge Base [2]
    - a small knowledge base of pharmacokinetic properties and drug-drug interactions

---

The proposed use case would go beyond the simple prototype

• Clinically relevant focus
  – Age-related pharmacokinetic changes and drug-drug interactions affecting antidepressants, antipsychotics, and sedative hypnotics

• State-of-the-art methods
  – Computer assisted scientific discourse mark-up
  – Natural language processing to identify potentially relevant claims
  – Intensive use of linked data sources

• Deliverables:
  – Software: A system that provides pharmacokinetic information on age-related clearance changes, metabolic clearance pathways, and pharmacokinetic drug-drug interactions for psychotropic, antidepressant, and sedative hypnotic drugs that are currently marketed in the United States - both from the existing label and newly extracted facts from the literature.
  – A W3C practice note on how to deploy the system for any given drug, at least one journal article describing significant results from developing the system deliverable and results from user testing.
  – ...
Success criteria

• Success Criteria
  – Ability to generate mashups for a significant number of PIs that provide complete and up-to-date content on drug metabolic pathways and DDIs.
  – Ease of replication for other drug PIs
  – Possible test integration into the Structured Product Labeling

• Other desirable outcomes
  – Improved state-of-the-art in claim identification (NLP).
  – Improved state-of-the-art in scientific discourse modeling
Partners

• Drug Informatics partners:
  – Richard Boyce, University of Pittsburgh

• Possible drug information partners:
  – Elsevier data
  – Others?

• Content providers:
  - MEDLINE abstracts in PubMed
  - XML-formatted full text articles in PubMedCentral
  - FDA-approved drug PIs in Structured Product Label format from DailyMed
  - Elsevier books, databases and journals

• Text mining collaborators:
  – Henk Harkema, University of Pittsburgh
  – Maria Liakata, EBI

• Scientific Discourse ontology specialist:
  – Jodi Schneider
  – Anita de Waard

• Possible pharma industry partners:
  – ?
• Questions and comments?