QUARTERLY REPORT – OCTOBER 2014

Dr. Rainer Winnenburg
rainer.winnenburg@nih.gov
NLM, Lister Hill National Center for Biomedical Communications, Cognitive Science Branch
Mentor: Dr. Olivier Bodenreider

Overall objectives

My research in the Medical Ontology Research Group focuses on various aspects of interoperability of biomedical data, including terminologies and ontologies, derived artifacts, such as value sets, and clinical datasets, using Semantic Web technologies. More specifically, I focus on the development of a framework and workflows based on terminology integration systems, such as the UMLS and RxNorm, to analyze and compare value sets used for clinical quality measures. Other applications include analyzing co-prescription data from the perspective of drug-drug interactions, alignment and comparison of drug information sources, and extraction and characterization of adverse drug events (ADE) from MEDLINE indexing.

Work since July 1, 2014

CDER Critical Path Project The aim of this collaboration between the Center for Drug Evaluation and Research (CDER) at the FDA and the NLM is to develop a novel analytic tool for quantitative drug-adverse event safety signal detection based on mining the biomedical literature.

I prepared another result set that contains updated drug safety information (covers currently all relevant articles before 2013 and > 85% of the articles published in 2013). I added provenance information for each drug – adverse event pair denoting how the relation was established (directly asserted, derived through parent concept, associated descriptor, or pharmacological action, etc.).

We are currently preparing a manuscript to be submitted to the Journal of Biomedical Informatics, in which we summarize the lessons we learnt while extracting the ADE information from MEDLINE articles leveraging MeSH descriptor / qualifier indexes. We also contributed to a poster that was presented at the 54th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC 2014), in which we demonstrate how we detect drug-adverse event safety signals through quantitative data mining of MEDLINE indexing terms based on a pilot study on fluoroquinolones. In another paper that was submitted to Nature Scientific Data, we contributed to the development and evaluation of a time-indexed reference standard of adverse drug reactions.

Previously, we had decided to look into the possibility of extracting adverse events directly from ClinicalTrials.gov and contrasting these findings to the signals obtained from the literature. I worked on several proof-of-concept studies demonstrating the feasibility of automatically extracting adverse events from ClinicalTrials.gov. On a set of 12k relevant clinical trial records (downloaded as XML), which are classified as completed and have results, I developed methods to (1) identify drugs from the intervention section of a clinical trial and map the active ingredient(s) to RxNorm CUIs, (2) detect these drugs in each column of the *Serious* and *Other Adverse Events* tables, and (3) extract quantitative and qualitative safety information for these drugs (or for different doses of the same drug). It is planned to explore the systematic extraction of ADEs from ClinicalTrials.gov beyond this preliminary work in a new inter-agency agreement between NLM and the FDA in the near future.

Exploring adverse drug events at the class level While the association between a drug and an adverse event (ADE) is generally detected at the level of individual drugs (e.g., between aspirin and Reye syndrome), it is often useful to discuss ADEs at the class level, i.e., at the level of pharmacologic classes (e.g., the ototoxicity of aminoglycosides). Some ADEs can be observed with every individual drug in a class, whereas some ADEs are associated with some class members, but not with all of them. The objective of this work was to explore the contribution of individual drugs to the class signal. More specifically, we proposed a visual representation of the association between drug classes and ADEs, which forms the basis for identifying strong associations. Using a similar visualization technique, we leveraged the hierarchical relations in drug classifications and ADE ontologies to investigate these associations at increasing levels of resolution. Applied to the ADE pairs we retrieved from the CDER Critical Path Project (drugs mapped to ATC single drugs and classes), we were able to uncover known associations, e.g., between fluoroquinolones and tendon injuries, and between statins and rhabdomyolysis. The findings gained from our exploratory techniques should be of interest to the curators of ADE repositories and drug safety professionals. This work was accepted as a paper at the International Workshop on Vaccine and Drug Ontology Studies (VDOS 2014).

Eliciting the intension of medication value sets As part of the summer student project of Nathaniel Bahr, I helped supervise the adaptation of methods we had originally developed for assessing the completeness, correctness and non-redundancy of diagnosis value sets (SNOMED-CT, ICD 9 and 10), making them applicable for drug value sets composed of codes from RxNorm. In this summer student project, we elicit the intension of medication value sets used in the latest release of clinical quality measures (CQMs) in the Meaningful Use incentive program by comparing them systematically to drug classes from ATC, MeSH, DailyMed, and NDF-RT, identifying the collection of classes that best characterize each value set. The results will be presented by Nathaniel Bahr in his Medical Informatics summer student final presentation on October 17, 2014.

Annual International Conference on Intelligent Systems for Molecular Biology (ISMB 2014) I presented two peer-reviewed papers at two Special Interest Group meetings (SIGs) at the ISBM in Boston: one paper on *Evaluating the consistency of inferred drug-class membership relations in NDF-RT* at the SIG meeting "Bio-Ontologies" on July 11, and one paper on the *Coverage of Phenotypes in Standard Terminologies* at the SIG session "Phenotype Day" on July 12, 2014.

NLM Fellows' Seminar I presented parts of my research on the analysis of clinical value sets and drug safety data using semantic data integration techniques on August 22, 2014.

Meeting of the LHNCBC Board of Scientific Counselors I presented two recent research projects as part of the Medical Ontology Research group's report on *Drug Terminology and Ontology* at the 64th meeting of the LHNCBC Board of Scientific Counselors on September 18-19, 2014. The presentation was based on two previously published papers: *A framework for assessing the consistency of pharmacological classes across sources* (Integration) and *Evaluating the consistency of inferred drug-class membership relations in NDF-RT* (Quality assurance).

Next-step work

In the following six weeks I will finish my work on ongoing projects such as the *CDER Critical Path Project*, with the emphasis on documentation and dissemination. If time allows, we plan to resolve some aggregation / granularity issues we found for some of our mappings from MeSH terms we identified as manifestations to terms from the target terminology, MedDRA. I would also like to look further into the extraction of drug safety information from ClinicalTrials.gov and the FDA Adverse Event Reporting System (FAERS).

Publications

Peer-reviewed conference papers / journal publications

Rainer Winnenburg and Olivier Bodenreider. *Issues in creating and maintaining value sets for clinical quality measures*. **Presented** at the AMIA 2012 Annual Symposium and *AMIA Annu Symp Proc*. 2012;2012:988-96. Epub 2012 Nov 3.

Rainer Winnenburg and Olivier Bodenreider. *Exploring pharmacoepidemiologic groupings of drugs from a clinical perspective.* Presented at MedInfo 2013, 14th World Congress on Medical and Health Informatics, Copenhagen, Denmark.

Rainer Winnenburg and Olivier Bodenreider. *Metrics for assessing the quality of value sets in clinical quality measures*. Presented at AMIA 2013 Annual Symposium, Washington, DC.

Rainer Winnenburg, Laritza Rodriguez, Fiona Callaghan, Alfred Sorbello, Ana Szarfman, and Olivier Bodenreider. *Aligning Pharmacologic Classes Between MeSH and ATC.* Presented at the International Workshop on Vaccine and Drug Ontology Studies (VDOS-2013) in conjunction with the 4th International Conference on Biomedical Ontology, Montreal, Canada.

Rainer Winnenburg and Olivier Bodenreider. *A framework for assessing the consistency of pharmacological classes across sources*. Journal of Biomedical Semantics 2014:5:30

Rainer Winnenburg and Olivier Bodenreider. *Desiderata for an authoritative representation of MeSH in RDF*. AMIA Annu Symp Proc 2014:(in press).

Rainer Winnenburg and Olivier Bodenreider. *Evaluating the consistency of inferred drug-class membership relations in NDF-RT*. Proc 17th ISMB'2014 SIG meeting "Bio-Ontologies" 2014

Rainer Winnenburg and Olivier Bodenreider. *Coverage of Phenotypes in Standard Terminologies*. Joint Bio-Ontologies and BioLINK ISMB'2014 SIG session "Phenotype Day" 2014:41-44.

Rave Harpaz, David Odgers, Greg Gaskin, William DuMouchel, **Rainer Winnenburg**, Olivier Bodenreider, Anna Ripple, Ana Szarfman, Alfred Sorbello, Eric Horvitz, Ryen W. White, Nigam H. Shah. *A time-indexed reference standard of adverse drug reactions.* Nature Scientific Data 2014:(submitted).

Rainer Winnenburg and Olivier Bodenreider. *Exploring adverse drug events at the class level*. Proceedings of the International Workshop on Vaccine and Drug Ontology Studies (VDOS 2014) 2014:(in press).

Posters

Rainer Winnenburg and Olivier Bodenreider. *Mapping drug entities between the European and American standards, ATC and RxNorm.* Data Integration in the Life Sciences, DILS 2012, College Park, MD, 06/28 – 29, 2012.

Rainer Winnenburg and Olivier Bodenreider. *Issues in Creating and Maintaining Value Sets for Clinical Quality Measures.* 26th Annual Research Festival, NIH, 10/09 - 12, 2012.

Rainer Winnenburg and Olivier Bodenreider. *Metrics for assessing the Quality of Value Sets in Clinical Quality Measures.* 27th Annual Research Festival, NIH, 11/06 - 08, 2013.

Steven Emrick, Duc Nguyen, Pishing Chiang, Philip Chuang, Maureen Madden, **Rainer Winnenburg**, Robert C. McClure, Ivor D'Souza, and Olivier Bodenreider. *The NLM Value Set Authority Center at 1 year.* AMIA 2013 Annual Symposium, Washington, DC, 11/18 2013.

Maureen Madden, Pishing Chiang, Emir Khatipov, Philip Chuang, Duc Nguyen, Ivor D'Souza, **Rainer Winnenburg**, Olivier Bodenreider, Julia Skapik, Robert C. McClure, and Steve Emrick. *Creating, maintaining and publishing value sets in the VSAC*. AMIA Annu Symp Proc 2014:(accepted).

Rainer Winnenburg and Olivier Bodenreider. *A Framework for Assessing the Consistency of Drug Classes across Sources.* 2014 NLM Informatics Training Conference, Pittsburgh, Pennsylvania, 06/17 - 18, 2014

Alfred Sorbello, Rave Harpaz, Ana Szarfman, Olivier Bodenreider, **Rainer Winnenburg**, Anna Ripple, Joseph Tonning, Henry Francis. *Detecting drug-adverse event safety signals through quantitative data mining of MEDLINE indexing terms: A pilot study.* Proceedings of the 54th Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC 2014) 2014.