

11-12-2011

# Implications of observation-fact modifiers to i2b2 ontologies

Jack W. London

*Kimmel Cancer Center, Thomas Jefferson University, Jack.London@jefferson.edu*

Devjani Chatterjee

*Kimmel Cancer Center, Thomas Jefferson University, Devjani.Chatterjee@jefferson.edu*

## Let us know how access to this document benefits you

Follow this and additional works at: <http://jdc.jefferson.edu/kimmelccfp>Part of the [Oncology Commons](#)

### Recommended Citation

London, Jack W. and Chatterjee, Devjani, "Implications of observation-fact modifiers to i2b2 ontologies" (2011). *Faculty papers Kimmel Cancer Center*. Paper 15.<http://jdc.jefferson.edu/kimmelccfp/15>

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Faculty papers Kimmel Cancer Center by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

As submitted to:  
*2011 IEEE International Conference on Bioinformatics and Biomedicine  
Workshops (BIBMW).*  
12-15 November 2011, Pages 929-930  
10.1109/BIBMW.2011.6112507

London, J. W., & Chatterjee, D. (2011). Implications of observation-fact modifiers to i2b2 ontologies. *2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, 12-15 November, 929-930.

Jack W. London, PhD and Devjani Chatterjee, PhD  
Kimmel Cancer Center, Thomas Jefferson University, Philadelphia PA  
jack.london@KimmelCancerCenter.org

### Abstract

*Biomedical translational research can be facilitated by integrating clinical and research data. In particular, study cohort identification and hypothesis generation is enabled by the mining of integrated clinical observations and research resources. The “informatics for integrating biology and the bedside,” or i2b2, framework is widely used for this biomedical data mining. The i2b2 “star schema” data model using entity-attribute-value (EAV) formatted concepts is a very efficient strategy for querying large amounts of data. However, until the most recent i2b2 release, the utility of the platform was somewhat constrained by the limitations on being able to express “facts about facts” – i.e., modify the observations about the patients. We have found that exploiting the new modifier functionality has significantly and favorably impacted the design of i2b2 ontologies, leading to easier and more meaningful query results.*

### Introduction

Research data marts are sources of clinical and research data on individuals for investigators. Clinical data including demographics, diagnoses, laboratory values, medications, procedures, and outcomes, can be extracted from clinical systems, transformed, and loaded into a clinical data warehouse (CDW). This CDW can in turn populate a research data mart (RDM) with a de-identified set of these data, in which HIPAA-defined Personal Health Information has been removed. Clinical data in the RDM can be augmented with research information, including annotation describing biospecimens available to researchers, clinical trial information, and registry data.

At Thomas Jefferson University we have deployed the i2b2 framework [1] to provide investigators with a simple to use query tool for cohort identification and hypothesis generation. Hospital clinical data (including demographics, diagnoses, procedures, medications, and laboratory results) were integrated with research biorepository annotation and cancer tumor registry data in a de-identified RDM [2]. We created ontologies for both the biospecimen and registry data. (*caTissue*, the NCI’s cancer Biomedical

Informatics Grid (caBIG<sup>®</sup>) biorepository management tool is used for research biobanking at TJU.) The ontologies we developed were constrained by the data model’s ability to easily express *facts about patients*, but not readily express *facts about facts*. For example, the observation that an individual had banked colon tissue could be straightforwardly represented, but adding other modifiers, such as that the tissue was frozen and malignant, was only possible by constructing redundant ontologies.

Release 1.6 of i2b2 has utilized an extension to the basic star schema model that allows for modifiers to be directly employed in the observation-fact dimension. This has allowed us to simplify the ontologies we use, particularly those for biospecimen annotation and registry data.

### Methods

The biospecimen and cancer registry ontologies that we developed for our initial deployment of the i2b2 web-based query application (version 1.5) were redesigned to utilize the modifier functionality available with i2b2 release 1.6. The extraction-transform-load software procedures that moved the data from their sources (*caTissue* and *Metriq*) to the i2b2 database were modified to support the new ontologies. User training was updated to use these new query strategies.

### Conclusion

The ability to directly represent observation-fact modifiers has significantly enhanced the investigators’ query capabilities, allowing them to more specifically describe the attributes they are seeking, while navigating a simpler ontology classification scheme. Of great consequence also is the ability to now include a HIPAA-compliant specimen identifier to the i2b2 specimen annotation, providing a means for an *honest broker* to easily order the material identified by the researcher’s query. The value of RDMs is greatly enhanced by observation-fact modifiers.

## References

[1] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). JAMIA 2010;17:124-130.

[2] London J.W., Housman D., Sagenich C., Chatterjee D., Jatkar T., Isikoff J.M., Palchuk M.B. Integration of Research Biospecimen Annotation with Clinical Data in an i2b2 Research Data Mart. American Medical Informatics Association Annual Symposium, Washington, DC, November 2010.