

# SSO: The Syndromic Surveillance Ontology

Okhmatovskaia A, Chapman WW, Collier N, Espino J, Buckeridge DL  
McGill Clinical and Health Informatics, McGill University

## OBJECTIVE

To create a software ontology that encodes agreement among experts about how Emergency Department (ED) chief complaints are grouped into syndromes of public health importance. The ontology must be straightforward, scalable, reference existing terminologies, and support automated surveillance.

## BACKGROUND

Most syndromic surveillance systems use data from ED visits, often free-text chief complaints [1]. Classification of chief complaints into syndromes is inconsistent, however, due to the lack of agreement about the concepts that define a syndrome and how individual terms or strings map to these concepts. Two threads of research have made it possible to address this problem now. First, an effort supported by the ISDS to develop a set of consensus definitions for syndromes in terms of ED chief complaints [2]. Second, experience in using ontology to encode standards and drive knowledge-based surveillance systems, most notably: to support surveillance of media [3], to enhance free-text classification [1], and to support modular aberrancy detection [4]. Uniting these two threads is critical to ensure that consensus syndrome definitions are disseminated broadly, maintained collaboratively, and incorporated easily into automated systems.

## METHODS

We used a set of existing consensus syndrome definitions [2] as the conceptual foundation for our ontology. These definitions were generated by an expert group of syndromic surveillance developers and practitioners for four syndromes: Influenza-Like Illness, Constitutional, Respiratory, and Gastrointestinal, with Respiratory and Gastrointestinal syndromes having sensitive and specific definitions. The definitions were structured in a way that: (a) explicitly defines a syndrome through a list of concepts that map to that syndrome, (b) explicitly defines a concept through related and synonymous concepts, (c) maps concepts to existing standardized vocabularies, and (d) allows for easy modification of definitions as surveillance practice evolves. We encoded this structure using Protégé-OWL to create the Syndromic Surveillance Ontology (SSO). OWL (Web Ontology Language) is the current standard for ontology development, specifically designed to facilitate machine interpretability, which is crucial for using the SSO to drive automated surveillance systems.

## RESULTS

The SSO contains four main classes that define fundamental concepts about clinical syndromes used for public health surveillance: Syndrome, Clinical Condition (member of sensitive or specific definition of a syndrome), Clinical Concept (used to define Clinical Conditions) and, Coding (links Clinical Conditions to major external coding systems, such as SNOMED). Individual syndromes, clinical conditions, concepts and coding systems are represented as the subclasses of the four main classes. Relationships among the classes are encoded using OWL class properties. For example, one of the properties of Syndrome class is *hasSensitiveDefinition*, which can be filled in by multiple Clinical Conditions. Clinical Condition, in turn, has such properties as *hasRelatedConcepts* and *hasExternalCodes*. The resulting entities encode highly specific knowledge (e.g. keywords defining a clinical concept), which makes the SSO readily useable in an automated surveillance system. The SSO is freely available online through BioPortal at <http://bioportal.bioontology.org>.

## CONCLUSIONS

The SSO will support the consistent and transparent application of consensus syndrome definitions to a number of surveillance activities. It provides a computable repository of knowledge that can be used to drive a chief complaint classifier and ensure that the classification represents the experience of surveillance systems across the country. The SSO also encodes standard definitions in a transparent and shareable format that can be used to compare the performance of different chief complaint classifiers and enable sharing of classification knowledge across surveillance systems.

## REFERENCES

- [1] Lu H-M, Zeng D, Trujillo L, Komatsu K, Chen H. Ontology-enhanced automatic chief complaint classification for syndromic surveillance. *Journal of Biomedical Informatics*. 2008;41(2):340-56.
- [2] Chapman WW, Christensen LM, Wagner MM, et al. Classifying free-text triage chief complaints into syndromic categories with natural language processing. *Artificial Intelligence in Medicine*. 2005;33(1):31-40.
- [3] Collier N, Kawazoe A, Jin L, et al. A multilingual ontology for infectious disease outbreak surveillance: rationale, design and challenges". *Journal of Language Resources and Evaluation*. 2007;40(3-4):405-13.
- [4] Buckeridge DL, Okhmatovskaia A, Tu S, O'Connor M, Nyulas C, Musen MA. Understanding detection performance in public health surveillance: Modeling aberrancy-detection algorithms. *Journal of the American Medical Informatics Association* 2008 Nov-Dec;15(6):760-9.