Reviewer's report

Title: PHSkb: A Knowledgebase to Support Notifiable Disease Surveillance

Version: 1 Date: 29 March 2005

Reviewer: AJ Valleron

Reviewer's report:

The authors underline that notifiable disease surveillance has still poor timing, low sensitivity, and that the surveillance protocols and definitions vary with state directives, and often change with time. There is no standardized vocabulary, nor uniform policy on case definitions, reporting criteria, etc.

This is the rationale for this paper. It proposes the use of Protégé, a reference software from Stanford. Protégé shows two different characteristics: it is an ontology editor (where classes, class hierarchy and relationships between classes and properties of these relationships must be chosen and defined), and a tool facilitating data entry through customized data entry forms.

The main contribution of Protege is the ontological one which requires that users model the knowledge used in the field under study (here: public health surveillance).

This paper addresses different problems:
- the need for standardized vocabularies and terminologies in surveillance.
- the high variability between states in terms of epidemiological surveillance (diseases, methods, etc.)
- the hypothetical link between an improved data and knowledge representation, and the quality of surveillance (expressed in terms of timeliness, sensitivity, etc.)
- the modeling of the disease surveillance system.

I think that the paper suffers from this disparity of questions, and should be centered more on the use of an ontology to model the disease surveillance.

For example, Tables 1 to 6 are descriptive information that one would expect to find on some official public health Website rather than in a scientific journal (I was surprised: Is Table 2 « original »? It looks as a very basic description of the state surveillance that should be on the CDC website?).

Regarding the surveillance ontology, which should be (in my opinion) the focus of this paper, I find what is shown not convincing. All the results are expressed on Figure 1 that displays which classes, relations, instances were chosen. For instance in the class « Infectious agent » and to my surprise, 6 instances are given (name, SNOMED code, Bioterror agent, associated disorder, jurisdiction reportable in). Rather I would expect subclasses (such as viral, bacterial, parasitic diseases), or at least instances allowing to enter such concepts. By contrast, two instances deal with bioterrorism, which is an important problem but not to the point to be
2/6 of the instances of the class "infectious agent". Similarly, I would put "jurisdiction reportable in ( ?)" in the class "Jurisdiction" and make a link from Infectious agent to the class Jurisdiction. The same remarks apply to the other classes.

Note also that the entire paper deals only with the surveillance of infectious disease and not for instance with that of environmental diseases. This should be stated in the title, and abstract.

Note finally that PHSkb was not downloadable at the FTP site indicated (when I tried).

- **Major Compulsory Revisions**

  Focus on the ontology. Give an in depth description of the chosen knowledge representation, and discuss the choices made.

- **Minor Essential Revisions**

  Delete Tables 1 to 6 which must be placed in a public health bulletin, not in a public health informatics paper

**What next?:** Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

**Level of interest:** An article of importance in its field

**Statistical review:** No

**Declaration of competing interests:**

I declare that I have no competing interests