Model-driven support for a vaccine study in Kathmandu

Jim Davies, Jeremy Gibbons, Steve Harris,
Oxford University Computing Laboratory

Jane Metz, Andrew J. Pollard, Matthew Snape
Oxford Vaccine Group, Department of Paediatrics, University of Oxford

Abstract

We report on the use of model-driven semantic technologies in support of a study into the effectiveness of paediatric vaccination against pneumococcal disease in Nepal.

1 Pneumococcal disease

Pneumonia and related diseases are a major public health problem. The World Health Organization estimates that diseases caused by the *Streptococcus pneumoniae* bacterium (*S. pneumoniae*, or pneumococcus)—including pneumonia, meningitis, and bacteraemia—cause 1.6 million deaths per year [1]; among these are the deaths of 0.7–1 million children aged under five years, 90% of whom live in developing countries (see Figure 1). Serious pneumococcal diseases are the number one vaccine-preventable cause of death in children under five.

There are 91 different serotypes of pneumococcus. The spectrum of prevailing serotypes varies over time and geography; however, common serotypes are consistently identified worldwide, and the 13 most common serotypes cause at least 70% of invasive pneumococcal disease in children [1].

The growing resistance of *S. pneumoniae* to commonly used antibiotics is a serious and rapidly growing problem; consequently, the prevention of pneumococcal disease through immunization is increasing in importance. A vaccine against 23 of these serotypes (the plain polysaccharide, 23-valent vaccine) is currently recommended for the elderly in the UK; however, this vaccine does not induce a robust immune response in young children. An alternative vaccine, the 7-valent pneumococcal conjugate vaccine (PCV7), is more immunogenic in children, and has been introduced into the routine childhood immunisation schedule of many industrialized countries in the last decade.

However, although PCV7 provides good serotype coverage in the West, it will not provide sufficient coverage for the different profiles throughout Asia and Africa. A recent study of pneumococcal disease among young children in Nepal [2] finds that only 6% of isolated specimens had a serotype covered by PCV7, and reports that "a new generation of pneumococcal vaccines that prevent infection with a broader range of serotypes may be necessary to most effectively control pneumococcal disease in young children in Kathmandu". Two such vaccines have been developed: PCV10, recently licensed for use in Europe, and PCV13, currently in the late stages of approval.

Nepal is one of the world's poorest countries, with a very high mortality rate for young children. It is one of the 72 developing countries eligible for financial assistance from the GAVI Alliance (Global Alliance for Vaccines and Immunisation), which is funded by five national governments and the Bill and Melinda Gates Foundation. In a collaboration between Patan Hospital (one of the main hospitals in Kathmandu), Otago University in New Zealand, and Oxford University, we are conducting a study [3] to determine the effectiveness of PCV10 vaccination in infants in Nepal—treating them with a variety of vaccination programmes, and measuring their serotype-specific antibodies afterwards. The study will be supported by software tools generated automatically from a model of the trial protocol, according to an approach we have been calling *semantic frameworks*.

2 Semantic frameworks

The *CancerGrid* project [4] set out in 2005 with funding from the UK Medical Research Council to improve information management for large-scale Phase III cancer clinical trials, with an emphasis upon the re-usability of the data obtained. The project employs a 'semantics-driven' approach:

- Information about the data to be collected in a study is made available in a standard, computable form. This semantic metadata will include details of the protocol, information about the context of the observation, the range of possible values, and their intended interpretation.
- This metadata is used to configure or generate the tools used in the acquisition and processing of information: typically, case report forms and software services. These tools will automatically establish and maintain the association between each item of data and its semantics.
- The recorded semantics can be published, with references to existing standards and datasets, to allow automatic identification of related research, opportunities for collaboration, and data re-use. Study descriptions, forms, and specific configurations for services can be published also, allowing re-use of design components.

The approach is intended to increase quality, and reduce cost, in individual studies, and also to increase value through the re-use or combination of data, and through the coordination of study designs. The reduction in cost arises partly from the reduction in bespoke development activity (the software required is derived automatically from the study model) and partly from the use of standard software components: in particular, those of Microsoft Office with SharePoint server.

The initial CancerGrid project ended in 2008, but work has since progressed in a number of directions. The *Accelerating Cancer Research Using Semantics-Driven Technology* project [5], funded by Microsoft Research, is extending the work done for Phase III trials into earlier-phase studies. The *Evolving Health Informatics* project [6], funded by the UK Engineering and Physical Sciences Research Council, is exploring the contribution of semantic frameworks to three scenarios in information driven health. One of these scenarios concerns rapid response to an emerging infectious disease pandemic, and entailed expansion of the scope of the CancerGrid approach to vaccinology.

The work on vaccine studies neatly complements the main theme of work on cancer, allowing technologies to be prototyped across the whole of the study lifecycle, within months rather than years, and provides a useful additional test of the applicability of the technologies outside the original domains of studies in breast and ovarian cancer. Moreover, the involvement with the developing world is a good match for our model-driven approach—a study based in Nepal cannot afford the luxury of a hand-built information system, so the automatic generation of tools from the study protocol is appealing—and resulting service-oriented architecture—rather than posting paper copies of case report forms, risking postal delays or losses, they are completed electronically in Kathmandu and uploaded to a server in Oxford.

The PCV10 study is undergoing ethical review and has received financial support, and is due to start soon. We are currently in the process of modelling the study in order to generate customized tools.

3 Acknowledgements

The Chief Investigator of the PCV10 study [3] is Dr Neelam Adhikari, Director of Paediatrics at Patan Hospital in Nepal. In addition to him and the listed authors, the work reported here is being conducted by Charles Crichton, Dominic Kelly, David Murdoch, Stephen Thorson, Andrew Tsui, and Shyam Upreti; their contribution is gratefully acknowledged. The Nepal study is funded by PneumoADIP (the Pneumococcal Accelerated Development and Introduction Plan) and the Hib Initiative at Johns Hopkins University, with a grant to the University of Oxford; PneumoADIP and the Hib Initiative are in turn funded in full by the GAVI Alliance (the Global Alliance for Vaccines and Immunisation). The continuing CancerGrid project is partially supported by a grant from Microsoft Research.

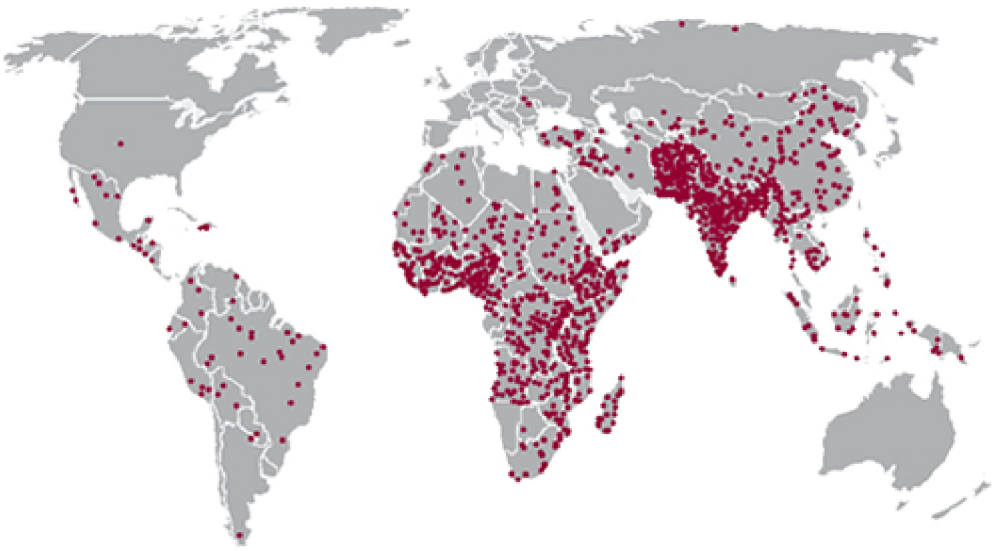

Figure 1: Pneumonia deaths in children under five (1 dot denotes 1000 deaths) [7]

References

[1] World Health Organization. Pneumococcal conjugate vaccine for childhood immunization. *Weekly Epidemiological Record*, 82(12):93–104, 23rd March 2007. WHO position paper.

[2] E. J. Williams, S. Thorson, M. Maskey, S. Mahat, M. Hamaluba, S. Dongol, A. M. Werno, B. K. Yadav, A. S. Shah, D. F. Kelly, N. Adhikari, A. J. Pollard, and D. R. Murdoch. Hospital-based surveillance of invasive pneumococcal disease among young children in urban Nepal. *Clinical Infectious Diseases*, 48:S114–S122, 2009.

[3] N. Adhikari, A. Pollard, D. Murdoch, S. Thorson, S. Upreti, D. Kelly, M. Snape, and J. Metz. A randomised open-labelled immunogenicity study of a 10 valent pneumococcal vaccine (PCV10) given as part of the routine infant immunisation schedule to children in Kathmandu, Nepal. Study protocol, Department of Paediatrics, University of Oxford, May 2009.

[4] J. Brenton, C. Caldas, J. Davies, S. Harris, and P. Maccallum. CancerGrid: Developing open standards for clinical cancer informatics. In *Proceedings of the UK e-science All Hands Meeting 2005*, pages 678–681, 2005. `http://www.allhands.org.uk/2005/proceedings/`.

[5] J. Brenton, J. Davies, J. Gibbons, and S. Harris. Accelerating cancer research using semantics-driven technology. In *Microsoft eScience Workshop*, Dec. 2008. `http://web.comlab.ox.ac.uk/publications/publication2850-abstract.html`.

[6] J. Davies, J. Gibbons, S. Harris, and D. Warzel. Evolving health informatics: Semantic frameworks and metadata-driven architectures. In *Microsoft eScience Workshop*, Dec. 2008. `http://www.comlab.ox.ac.uk/publications/publication2851-abstract.html`.

[7] PneumoADIP. Pneumococcal diseases. `http://www.preventpneumo.org/diseases/pneumococcal_diseases/`, last checked July 2009.